DOCUMENT RESUME

ED 237 731 CE 037 740

AUTHOR Rivera, William M.; And Others
TITLE Designing Studies of Extension Program Results: A Resource for Program Leaders and Specialists. Volume I--Text.
INSTITUTION Extension Service (DOA), Washington, D.C.; Maryland Univ., College Park. Cooperative Extension Service.; Maryland Univ., College Park. Dept. of Agriculture and Extension Education.
PUB DATE Oct 83
NOTE 122p.; For Volume II, see CE 037 741.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Adult Education; Adult Vocational Education; Agricultural Education; Community Development; Environmental Education; *Evaluation Methods; *Extension Education; Home Economics; Natural Resources; Postsecondary Education; Program Descriptions; *Program Evaluation; *Research Design; Secondary Education; Standards; *Summative Evaluation

ABSTRACT

This review of selected study designs, the first volume of a two-volume resource, is intended to advise state administrative program leaders and others with program evaluation responsibilities of how to design studies of cooperative extension programs. Three kinds of information are presented in the volume's three parts. The first part introduces the nature and purpose of studies of extension program results and places the design of studies into perspective. It reviews evaluation users, dimensions, general procedures, and standards. Alternatives to designing studies of extension program results are cited in part 2. Four basic study designs used in evaluating extension program results are examined ((1) survey (ex-post facto), (2) time-series, (3) comparison group, and (4) field experiment), and their relative suitability is linked to practical considerations based on the evaluation situation. This part also lists 10 specific facets of the study designs and illustrates how each may lead to alternative ways of constructing a design for a study. Part 3 provides a brief review of the conditions that lend themselves to alternative study designs and suggests contexts in which this resource publication should be viewed. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED237731

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W Rivera

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This two-volume publication was developed for the Cooperative Extension Service, University of Maryland, and the Extension Service, U.S. Department of Agriculture under a cooperative agreement. The Department of Agricultural and Extension Education, University of Maryland, assisted the Maryland Cooperative Extension Service in fulfilling its responsibilities in the cooperative agreement. The views expressed in this document are those of the authors, and do not necessarily reflect the positions or policies of the Maryland Cooperative Extension Service or the Extension Service, U.S. Department of Agriculture.

ACKNOWLEDGEMENTS

Many hours of concentrated effort were required to complete this publication. The authors are indebted to every person who contributed time and talents to its completion. Specifically, we would like to recognize:

The Extension Service, United States Department of Agriculture, (ES, USDA) for its support of this project including:
Mary Nell Greenwood, Administrator
John S. Bottum, Deputy Administrator

The Maryland Cooperative Extension Service (MCES), especially:
Craig Oliver, Director, MCES
Billy L. Coffindaffer, Associate Director, MCES

The University of Maryland, Department of Agricultural and Extension Education, particularly:
Clifford L. Nelson, Department Chairman
James W. Longest, Professor

A Maryland State advisory committee assisted in the selection of the abstracted study reports included in volume two of this publication. The committee's recommendations at the early stages of the project provided valuable insight and direction. We are grateful to:

Billy L. Coffindaffer, Associate Director, MCES
Clifford L. Nelson, Professor, University of Maryland
Mark A. Varner, Dairy Specialist, MCES
Karol Westelinck, Extension 4-H Program Coordinator, Prince Georges County, Maryland
Richard D. Wootton, Extension Director, Montgomery County, Maryland

An Interstate Advisory Board further assisted in the project; and was comprised of the following Cooperative Extension personnel: Charles D. Clark, Program Evaluation Specialist, University of Illinois; Oscar Strickland, Head of Program Development, Auburn University; Benjamin H. Weddle, Jr., Interim Director, University of New Hampshire; and Joan Wright, State Leader of Training, North Carolina State University. Members reviewed early drafts of the publication. Their advice and direction are most appreciated.

Reviews of an early draft were also provided by the following Cooperative Extension personnel:

Mary Andrews, Program Leader for Evaluation and Reporting (Michigan)
Milton Bauldauf, Program Leader, HEHN (ES, USDA)
Milton Boyce, Assistant Deputy Administrator, 4-H Youth, (ES, USDA)
Clarence J. Cunningham, Assistant Director (Ohio)
Laverne B. Forest, Leader, Program Development and Evaluation (Wisconsin)

Barbara Froke, Program Leader, Family Living and Nutrition (South Dakota)
Shirley H. Gerken, Evaluation Specialist (Virginia)
Ricardo Gomez, Staff Leader, Agricultural Programs (ES, USDA)
Tommie Lou Hunter, Program Leader, Home Economics (Virginia)
Howard E. Jones, Regional Director (Arizona)
Richard Krueger, Leader for Program Evaluation (Minnesota)
Howard Ladewig, Associate Professor of Rural Sociology (Texas)
George W. Mayeske, Program Evaluation Specialist, PDEMS (ES, USDA)
William Moyles, Assistant Statistician (California)
R. David Mustian, State Leader of Evaluation (North Carolina)
William Pietsch, Associate Director (North Dakota)
Midge Smith, Program Evaluation Specialist (Florida)
Bonnie Tanner, Program Analyst, HEHN (ES, USDA)
Joan S. Thomson, Coordinator for Staff Development (Pennsylvania)
Paul D. Warner, Associate Professor of Sociology (Kentucky)
Andrew J. Weber, Program Leader, Natural Resources (ES, USDA)
Phyllis E. Worden, Program Leader, Home Economics (Colorado)

A special thanks goes to those persons who were responsible for the printing of this resource, including Anne Pease, Steve Rothman, and Larry R. Whiting, Director, Information and Publications, MCES. We are particularly appreciative of the dedication and endurance shown by the project secretaries, Jackie Johnson and Judie McCaslin.

Finally, we are indebted to the authors of the abstracted study reports included in Volume II. Their work provides useful insights to all of us interested in examining the effectiveness of Extension programs in the United States.

William M. Rivera
Associate Professor
Department of Agricultural and Extension Education
University of Maryland

Claude F. Bennett
Program Analyst
Extension Service
U.S. Department of Agriculture

Sharon M. Walker
Research Assistant
Department of Agricultural and Extension Education
University of Maryland

CONTENTS

	Page
Acknowledgements	iii
Preface	vii
Part I: Background and Approach	
Chapter 1: Introduction	1
Purposes of the Resource	1
Guide to Using this Resource	4
Definitions	6
General Procedures	8
References	15
Chapter 2: The Place of this Resource in Extension Evaluation	17
Evaluation Users	19
Steps in Evaluation	21
Standards for Evaluation	25
Conclusion	26
References	27
Part II: Designs for Examining Results	
Chapter 3: Introduction to Study Designs	29
Overview of Selected Designs	29
The Survey Design	33
The Time-Series Design	38
The Comparison Group Design	39
The Field Experiment Design	41
References	45
Chapter 4: Options in Designing Studies	47
Selecting a Study Design	47
Overview of Facets of Study Design	50
References	63
Chapter 5: Study Designs, Facets and Examples	63
Survey Designs	65
Time-Series Designs	69
Comparison Group Designs	73
Field Experiment Designs	76
General Comments and a Reminder	79
Further Readings	81

Part III: Summary and Conclusions

Chapter 6: Summary 89

Decisions, Decisions 89
A Summary of Starting Points 90
Summary of the Text 99
References 101

Chapter 7: In Conclusion 103

Dimensions and Steps in Program Evaluation 103
Approaches and Methodologies 104
Past as Prologue 106
References 108

RECOMMENDED READINGS 109

APPENDICES 111

A. Procedures for Abstracting and Appraising Studies 113

B. Standards for Appraising Studies 117

C. Procedures for Obtaining Full Reports of Studies 121

PREFACE

For accountability and administrative decisions, there is increasing need for State Cooperative Extension program evaluation studies which are credible to State and Federal legislators and executives, university leaders and Extension administrators. This need is made apparent by the recently adopted Extension Accountability and Evaluation System. The system calls for the Extension Service, USDA, to provide staff development and technical assistance to State Extension Services' studies on the inputs, operations and impacts of Extension programs. In January 1983, Extension Service, USDA, entered into a cooperative agreement with the Maryland Cooperative Extension Service, through the University of Maryland's Department of Agricultural and Extension Education, to develop and publish a resource publication on designing studies to evaluate the results of Cooperative Extension programs.

The purpose of this two-volume resource is to advise state administrative program leaders and others with program evaluation responsibilities how to design studies of program results that are accurate and reliable. This resource may also aid in the development of more uniform standards for program evaluation within Cooperative Extension.

Volume I of the resource reviews selected study designs. These illustrate different approaches to examining the extent to which clientele behavior or status can be attributed to an Extension program. In order to show the feasibility of using a variety of evaluative study designs to examine program results, Volume II includes abstracts of 42 studies as examples of these study designs. The abstracts were selected

to represent Extension programs in agriculture, natural resources, community development, 4-H youth, home economics and overall county Extension programs. Each study exemplifies one or more of four basic study designs: (1) survey (ex-post facto), (2) time-series, (3) comparisor group, and (4) field experiment. These study designs and the example studies should not limit future evaluation efforts. They are intended to provide State Extension Services with background knowledge for designing studies applicable to their own evaluation interests and needs.

PART I - BACKGROUND AND APPROACH

CHAPTER 1: INTRODUCTION

Purposes of the Resource:

The publication of this resource is in partial response to three documents: (1) "Report of the National Task Force on Extension Accountability and Evaluation System" (1981), approved by the Extension Committee on Organization and Policy (ECOP); (2) "The Comptroller General's Report to the Congress on "Cooperative Extension Service's Mission and Federal Role Need Congressional Clarification" (1981); and (3) Program Evaluation in Extension, a 1980 report on the status of the organization for and practice of program evaluation by all State Extension Services.

The "Report of the National Task Force on Extension Accountability and Evaluation System" includes the recommendations that: (1) State Extension Services undertake indepth studies of the inputs, operations and impacts of selected programs, primarily in order to meet state or multi-state needs for accountability and evaluation; and (2) Extension Service, USDA, provide staff development and technical assistance to State Extension Services in the planning and conducting of such studies.

The General Accounting Office's "Cooperative Extension Service's Mission and Federal Role Need Congressional Clarification" recommends that the Secretary of Agriculture assume leadership, with ECOP, for developing and implementing a uniform evaluation system for Cooperative Extension, including clearly defined evaluation standards. This

resource uses a number of technical standards for conducting studies on Extension program results. These ad hoc standards may lead to further discussion within Extension regarding appropriate standards for evaluation studies within the Extension Accountability and Evaluation System.

Finally, the 1980 national study conducted by West Virginia University reports that: (1) 60 percent of the State Extension Services had not assigned specific responsibility for evaluation to any particular staff member; (2) almost two-thirds of state program leaders and almost three-fourths of county agents considered themselves prepared only to "some extent" or to a "small extent" to adequately plan; implement, or report program evaluations; and (3) 85 percent of State Extension directors feel similarly about their staffs' inadequate preparation to corduct program evaluations.

The majority of State Extension Services have no history of conducting studies of program results with the rigor expected by decision-makers (or, more likely, their staffs) at State and Federal levels. Typically, states do not have evaluation specialists to stimulate, conduct or give advice on studies of program results. Thus, the Extension Service, USDA, entered into a cooperative agreement with the Maryland Cooperative Extension Service to produce a resource publication that will assist program leaders and specialists in designing studies of program results which will be credible to study users in the Cooperative Extension Service and elsewhere.

The role and evaluative expertise of Extension program leaders and specialists will vary from study to study. However, this resource

publication should help program leaders and specialists perform the following roles with respect to studies of Extension program results:

- Advise Extension evaluation specialists as they design studies of program results.
- Serve on steering committees where program and evaluation personnel have joint responsibility for designing and conducting a study of program results.
- Formulate and manage contracts or cooperative agreements, with professional evaluators outside Extension, to study results of Extension programs.
- Design and conduct studies of program results in consultation with Extension or non-Extension evaluation specialists.

This resource publication contains examples of previous studies which demonstrate feasible ways to ascertain results of Extension programs using four selected study designs--the survey, time-series, comparison group and field experiment. The resource may also prove of value to evaluation specialists: (1) as a reference to previous Extension studies; (2) for inservice training and classroom education; and (3) for systematically reviewing choices in designing a study of Extension program results. Evaluation specialists may gain further appreciation of the optional methods demonstrated by these past studies and how these may be applied, at least in part, to current evaluation efforts. Moreover, the resource volumes should enhance communication among Extension program staff, university academic staff, and non-university evaluators.

Guide to Using this Resource

We intend for DESIGNING STUDIES OF EXTENSION PROGRAM RESULTS to provide three kinds of information. These are presented in the three parts of Volume I.

Part I introduces the nature and purpose of studies of Extension program results and places the design of studies into perspective. Part I reviews evaluation users, dimensions, general procedures and standards.

Part II cites alternatives to designing studies of Extension program results. These results may be (a) educational (knowledge, attitudes, skills and aspirations); (b) practices based on the use of new learning, or (c) impacts--the eventual economic, social or environmental results of new learning and practices. Part II examines four basic study designs used in evaluating Extension program results, and then links their relative suitability to practical considerations based on the evaluation situation. This part also lists ten specific facets of the study designs and illustrates how each of these facets may lead to alternative ways of constructing a design for a study. For consultation, the facets are referenced to the abstracts of previous studies of program results in Volume II. These abstracted studies are simply examples of the use of the four basic study designs--not necessarily exemplary studies.

Part III provides a brief review of the conditions which lend themselves to alternate study designs, and suggests contexts in which the resource publication should be viewed.

For ease in comparing the Volume I text on design principles with actual approaches and methods, abstracts of a selection of previous studies of programs are in Volume II of the publication. A review of the example studies may be useful in approaching one's own program evaluation responsibilities, for the abstracts show how others in the past have approached and attempted to answer study design questions. Thus, the user may wish to open Volume II to studies of immediate interest while reviewing principles of design formulation presented in Volume I. Users should note in particular whether the abstracts illustrate educational results, practices, long-term results or other program-related results -- such as clientele's use and rating of information received and the sources of such information.

The publication may be read sequentially, i.e., piece by piece. Or, users may wish to skip directly to Part II and/or Part III in order to obtain specific ideas for designing studies. Part I is not essential for selecting study designs or design facets. Our intention is that the publication should serve primarily as a resource.

References to other sources of published information useful in designing studies appear at chapter ends and at the close of Part II.

Since the resource is oriented towards a readership with varying degrees of expertise in planning and implementing studies of Extension program results, it contains some elementary material and may occasionally seem repetitive or redundant to those who are more experienced in program evaluation.

Whether a study measures immediate program outputs, follow-up practices or impacts, the question of its design is paramount. Design is a core consideration in choosing a methodology, influenced by the

intentions, procedures, timing and budget for the study. Only four designs are put forward in this resource. They are not the only designs for studies of program results, as is repeatedly stated throughout this resource. The designs advanced are not new, but are well-known; indeed, this resource brings them together partly because they tend to be popular among program evaluators.

The special contributions of this resource to the subject of study designs are: (1) it links overviews of the four designs and alternative facets within these designs to their application in previous studies of Extension program results, as a way of illuminating potential methods as well as pitfalls in future studies, and; (2) it systematically presents several Extension programming and evaluation situations which affect the choice of study design. Moreover, the publication addresses practical as well as ideal scientific considerations in the selection of study design.

Definitions

The term "program impact" is used variously. Sanders (1982) defines program impact as broadly as we define program results.

> Impacts may be intended or unintended, may be positive, negative or neutral in value, may be stable or unstable and fleeting; may be seen at the immediate closing of the program or service and/or may be seen a long time following the program or service, may appear for primary recipients and/or tertiary recipients, twice removed from being directly involved in the program or service.

Guidelines for the Extension Accountability/Evaluation (A&E) System (1983) define "Impact Studies" as technically valid indepth studies to assess: (a) the economic or social consequences of Extension efforts, or (b) other aspects of Extension inputs, operations or programs. Thus,

the Extension A&E System emphasizes studies of the economic and social consequences of Extension programs while allowing for other types of technically valid studies as noted above. The stress on economic and social consequences of Extension programs appears to be consistent with the request of the U.S. Congress (1977) that the Secretary of Agriculture provide an evaluation of these consequences. However, even though the emphasis may be on program consequences, the Extension A&E System allows for other types of valid studies, such as those mentioned above in (b).

Accordingly, we define program impact as the economic, social environmental and individual consequences (results) of program-induced learning and practices. These consequences or end results (Bennett, 1979 and 1980) emphasize the prevention, checking, reduction or solution of problems encountered by clientele. Ideally, evidence of program impact is expressed in terms of whether desired end results occurred, plus detection of any side-effects. However, without assurance that clientele have attained a certain level of performance or practice through program-induced learning, attributing desired end results to an Extension program may be meaningless.

Because direct measures of program impact are difficult to obtain, it has been suggested that utilization (e.g., clientele practices) be used as indirect or "proxy" indicators of impact (Wholey, et al. 1970; UNESCO, 1979). An impact study could be viewed as a two-way street with measurements of participant learning and performance or practice pointing toward impacts, and any measurements of end results (the impacts themselves) pointing back toward the original influences (including the program) which produced impact.

In summary, the Extension A&E System includes both broad and narrow definitions of "impact study." The broad definition of impact studies includes indepth studies of Extension inputs, operations and impacts. Impact study is here capitalized in reference to the full definition contained in the guidelines for the Extension A&E System (1983). The narrower definition of impact studies - studies of the economic and social consequences of Extension programs - appears to emphasize the end results of the education and practice induced by Extension programs. The narrow definition is a component of the broader definition.

Our conclusion is this: an impact study should somehow assess a program's final consequences: (a) preferably through providing evidence bearing directly on the program's end results, or (b) by discussing how a program's measured educational and/or practice results might be expected to produce its end results.

General Procedures

The 42 abstracted studies containing findings on Extension program results were selected by the authors from a pool of 153 studies conducted from 1961-1982. This pool was comprised of: (a) 148 studies of Extension program effectiveness selected (in 1978) by two social science research firms under contract to Extension Service, USDA (Bennett, 1980); and, (b) five additional studies selected from responses to an Extension Service, USDA, request for 1979-82 state studies on Extension program impacts. The USDA co-author (with the advice and assistance of other evaluation staff members of PDEMS, Extension Service, USDA) selected these five studies, using the criteria

and procedures adopted by the lead social science research firm. The research firms had been contracted to select studies with adequate methodological substantiation of findings and conclusions regarding Extension program results (Appendix A).

Procedures adopted for the present project included the following:

Task I: Inventory each of the 153 abstracts and classify the studies according to the following four study designs: survey, time-series, comparison group, and field experiment.

Task II: Review and appraise each of the abstracts according to the criteria set forth in the Extension Service, USDA, request for state studies from the period 1979-82 (Appendix B).

Task III: Select from within each of the designs several studies that exemplify several methods for gathering evidence of program results in each Extension program area.

Task IV: Review existing guidelines and standards for evaluations and relate these to studies of Extension program results.

Task V: Develop guidelines for selecting study designs and conducting studies that examine program results.

Task VI: Identify facets and options in implementing each study design, citing examples among the abstracts.

Task VII: Identify issues in design methodology based on appraisals of the methodologies of the 42 examples.

Task VIII: Draw conclusions and implications for designing Extension program results, including conditions under which each study design may be selected.

Task IX: Prepare a publication to aid state program leaders and specialists in their various roles in designing studies of program results, with the assistance of the interstate advisory committee and other State and Federal Extension personnel.

The 153 studies are classified among the four study designs as follows:

Survey	69
Time-series	42
Comparison group	31
Field experiment	11

Each study was then categorized by program area: agriculture, community development/natural resources, 4-H, home economics, or overall county Extension program. The distribution of studies by program area according to type of design is presented in Table I.

TABLE I Abstracted Studies (from Which Examples Were Selected) Grouped by Extension Program Areas and Study Design

	Extension Program Area					
Study Design	Agriculture	Com.Dev.& Nat.Res.	4-H Youth	Home Economics	Total Program	Total
Survey	20	12	6	23	8	69
Time-Series	12	2	11	17	0	42
Comparison Group	9	3	11	8	0	31
Field Experiment	3	0	2	6	0	11
Total	44	17	30	54	8	153

The 153 abstracts were reviewed by the authors and a Maryland State advisory committee of University of Maryland staff and State and local Extension personnel. From these abstracts, 42 were selected to represent a variety of methods for obtaining evidence on program results. The studies selected represent both those which required large study resources and those requiring less resources. Table II presents the number of studies selected from each study design according to the Extension program areas. It was sometimes difficult to place a given abstract within one design and program area category.

TABLE II Abstracted Studies Selected as Examples, Grouped by Extension Program Areas and Study Design

Study Design	Extension Program Area: Agriculture	Com.Dev.& Nat.Res.	4-H Youth	Home Economics	Total Program	Total
Survey	2	5	3	6	1	17
Time-Series	5	0	1	4	0	10
Comparison Group	3	1	3	1	0	8
Field Experiment	2	0	1	4	0	7
Total	12	6	8	15	1	42

Table III presents the names of the authors of the studies selected by study design and by program areas.

Table III: Authors of Abstracted Studies With Year of Publication Grouped by Extension Program Areas and Study Designs

	Agriculture	Community Dev. & Natural Resources	4-H Youth	Home Economics	Overall County Extension Program
Survey	Awa, NE & Van Crowder, R '77 Dardeau, JL '77	Brown, FL '67 Forest, LB et al. '76 Forest, LB '77 Frye, L & Miller, DR '76 Tait, JL '69	Rockwell, KS et al. '82 Steele, SM & Everson, N '78 Summers, JC & Zeller, FA '77	Bogenschneider, K '77 Goetting, MA '82 Glass, PL & Reese, BR '76 Kanarek, L '78 Williams, MJ '78 White, L & Ladewig, H '79	Forest, LB & Marshall, MG '77
Time-Series	Balliet, L '78 Carter, CE '80 Hartman, JA & Brown, EJ '70 Ladewig, HW & Edmondson, VW '72 Strickland, CL et al. '76		Marks, RJ '71	Brown, JS & Nelson, HY'77 Crowe, ND et al. '75 Kingdon, LB & Toensmeyer, UC '75 Texas Ag. Ext. Ser. '74	
Comparison Group	Aldrich, RJ '75 Alexander, FD & Longest, JW '62 Gruenhagen, RH '69	Street, P '73	Boone, EJ & White, E. '76 Mauk, WS '75 Subaima, GV '61	Green, LW et al. '72	
Field Experiment	Scherer, CW '77 Warren, RD et al. '66		Trent, C & Glass, JC '77	Bowering, J et al. '78 Honnold, R et al. '81 Trent, C & Kinlaw, R. '76 Yerka, B '74	

All of the 42 studies contain sufficiently strong evidence for determining the results of Extension programs. However, these 42 studies are not necessarily the "best" studies of Extension program results, methodologically speaking. First, the pool of studies from which the 42 were selected does not include all the methodologically adequate studies of Extension program results which were conducted between 1961-1982. For example, the 42 studies represent only quantitative studies, as no reviews of qualitative studies of Extension program results were available. Secondly, the 42 studies were selected from among others of perhaps equal technical quality in an effort to portray studies from different methodological approaches, different program areas and different geographic areas of the United States. Finally, examples were cited which might affect future studies relying on varying levels of resource availability. (Nearly one-half of the 42 studies were funded at least partially by Extension Service, USDA, or private sector special funds.)

Refer to the Table of Contents in Volume II for abstracts of individual studies listed according to the four study designs. Studies are ordered within a design category by alphabetical order of the author's name. Or, individual studies may be located by consulting the Index of Studies according to program area--where abstracted studies are listed alphabetically (by author's name) within program area categories. Finally, Volume II includes an index of studies by authors' alphabetized names.

The abstracts included in Volume II were edited but are not altered in substance from their sources in the research contractor publications. We chose to delete the study appraisals by the research contractors.

However, we do incorporate design-related issues within these appraisals into the discussions of each study design.

Copies of the complete reports of the abstracted studies in Volume II and of the studies referenced in the further readings are available upon request to the National Agricultural Library, Beltsville, MD 20705. The studies are referenced in the National Agricultural Library's computerized retrieval system. Appendix C provides specific instructions for obtaining full reports of the studies of Extension program results which are referenced in this publication.

References

Bennett, C. F. (1979) Analyzing impacts of Extension programs. Extension Service Circular-575. Washington, DC: U.S. Department of Agriculture.

Bennett, C. F. (1980). Teaching materials on 'seven levels of evidence'; A guide to Extension workers. Extension Service Circular-575, Supplement 1. Washington, DC: U.S. Department of Agriculture.

Bennett, C. F. (1980). Overview of contractual work to review, appraise and summarize studies of Extension program effectiveness, 1961-1978, with implications for future Extension program evaluation and accountability. Washington, DC: U.S.D.A. Extension Service.

Cronbach, L. J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

Comptroller General of the United States. (1981). Report to the Congress of the United States on Cooperative Extension Service's mission and federal role need congressional clarification. (CED-81-119). Washington, DC: United States General Accounting Office.

Extension Accountability/Evaluation System (1983). State Extension plan of work and report guidelines: October 1, 1983 - September 30, 1987. Washington, DC: Extension Service, U.S. Department of Agriculture.

Kappa Systems, Inc. (1979). Classification of selected Extension studies with impact findings. Vol I. Arlington, VA: U.S.D.A. Science and Education (under contract).

Kappa Systems, Inc. (1979). Extension program impact findings from selected studies conducted from 1961 to 1978. Vol. II. Arlington, VA: Author. Appendix to Vol.II, Supporting abstracts.

Kappa Systems, Inc. (1979). Guidelines for improving Extension program impact studies, Vol. III. Arlington, VA: Author.

National Association of State Universities and Land Grant Colleges. (1981). Report of the national task force on Extension accountability and evaluation system.

Sanders, J. (1982). Criteria for technically valid impact studies. Evaluation in Cooperative Extension programs: Proceedings from mini-session at evaluation network annual conference. Baltimore, MD., October 28, 1982, compiled by M.F. Smith (pp. 39-44). Gainesville, FL: Cooperative Extension Service.

Science Management Corporation. (1979). Inventory and classification of reports abstracted, Vol. I. Washington, DC: Author.

Science Management Corporation. (1979). Final Report: A review and appraisal of studies of Extension program effectiveness. Vol. II. Washington, DC: Author.

Summers, J. C., Miller, R. W., Carter, C. E., Young, R. E., Dempsey-Foss, C., & Beaumont, L. (1981). Program evaluation in Extension, Vol. I Findings and recommendations. Morgantown, WV: Center for Extension and Continuing Education, Office of Research and Development.

United Nations Educational Scientific and Cultural Organization (UNESCO). (1979). Evaluation of technical cooperation projects in education. Guide for the non-specialist practitioner. Draft. Paris: Author.

United States Congress. (1977). The Food and Agriculture Act of 1977, P.L. 95-113, Title XIV--National Agricultural Research, Extension and Teaching Policy Act. Washington, DC:9th Congress.

Wholey, J. S., Scanlon, J.W., Duffy, H.G., Fukumoto, J.S., & Voght, L.M. (1970). Federal evaluation policy: An overview. Washington, DC: Urban Institute.

CHAPTER 2: THE PLACE OF THIS RESOURCE IN EXTENSION EVALUATION

Traditional models of programming tend to portray evaluation as one phase of a cyclical process--the phase following program planning and program implementation.

However, recent models suggest that evaluation is a part of each phase of programming. Different authors divide the overall programming process into different phases, or title the phases differently. However, all of the models deal with the questions of deciding what kind of program to have, then how to conduct it, and finally [illegible] improvements.

For example, Stufflebeam (1971) identifies evaluations which help to:

- Select program purposes regarding intended clientele benefits (evaluations of program context)
- Select program designs capable of achieving the desired clientele benefits (evaluations of program inputs)
- Improve program designs or their implementation (evaluations of program process)
- Improve program designs or purposes in order to better achieve intended clientele benefits (evaluations of program products).

Some authors subdivide the four dimensions referred to by Stufflebeam, or attach different names to these dimensions. For example, input evaluations may be called formative evaluation; process evaluations may be called program monitoring. Relatedly, Dave (1980) divides evaluations of program products (results) into those which

examine short-term program results and those which examine longer-term results.

Our resource publication focuses on studies which help to evaluate program products, outcomes or results. Three levels of program results are defined and incorporated in this resource:

- Educational results: these include changes in clientele knowledge, attitudes, skills and aspirations.
- Practice results: these include clientele patterns of behavior, actions or performance stemming from educational results.
- End results: these include the consequences or impacts of program-induced educational and/or practice results.

Figure 2.1 shows that these three levels of program results are levels 5, 6, and 7 in a levels of evidence model (Bennett, 1979).

Figure 2.1

7. End Results

6. Practice Change

5. KASA Change (knowledge, attitudes, skills, aspirations)

4. Reactions

3. People Involvement

2. Activities

1. Inputs

"Value of information" studies also receive attention in this resource. These studies are closely related to studies of Extension program results, and are based on participants' ratings of the value of information they received from Extension, and from other sources for decisionmaking in some field of endeavor. Such evidence may imply the existence of program results although the nature of such results is not specified.

Our resource publication provides an overview of four main study schema for selecting, collecting and using evidence in evaluating a program; i.e., four study designs. These designs are: survey, time-series, comparison group and field experiment.

Evaluation Users

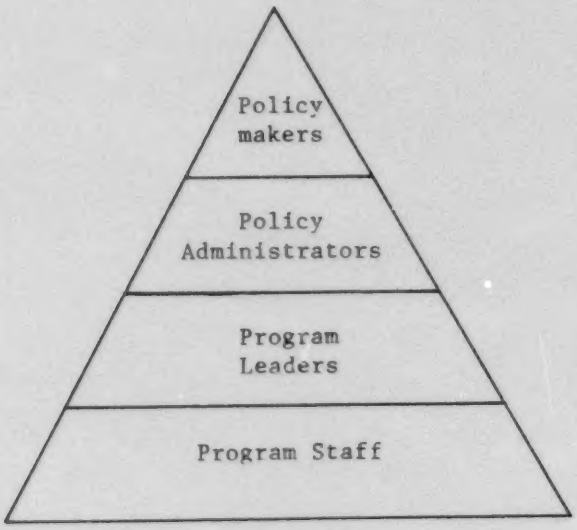
The question of who will use the evaluations of program results is central. Schmidt (1977) distinguishes between policymaker users and program manager (leader) users. The various types of program results are differentially germane to different kinds of users. Different kinds of users may use information on particular levels of program results, and not use information on other levels because of their differing responsibilities.

To illustrate this point, Rivera (1982) constructs a pyramid composed of program evaluation users from the top down (Figure 2.2):

- Policymakers (e.g. legislators and state and federal executives) who oversee policy administrators;
- Policy administrators for the program (e.g. Extension directors)
- Program managers (e.g., Extension program leaders)
- Program staff (e.g., Extension specialists and agents).

Figure 2.2

A Pyramid of Evaluation Users



As a general rule, the higher the position of users in the pyramid, the greater their need for information on a program's end results (impacts). Users high in the pyramid have less need for information on practice results and even less need for information on educational results.

Conversely, the lower the position of users in the pyramid, the greater their need for information on educational and practice results. Users lower in the pyramid have less immediate need for information on end results.

Generally speaking, policymakers and policy administrators are more in need of evidence regarding end results because this is more useful for making policy decisions on whether to initiate and/or continue a given program, and how many resources to commit to it. Program leaders and program staff are more in need of evidence on educational and

practice results because these are more useful for making judgments regarding the extent to which a program is exhibiting effective and efficient progress toward reaching the desired end results. It should be clarified that our resource publication focuses on a variety of ways to obtain evidence on program results, and we only touch here on the extremely important topic of uses of information regarding program results.

In considering different designs for identifying program results, it is important to select or construct a design that will provide intended study users with evidence of program results in which they are most interested. Of the 42 abstracted previous studies containing Extension program results, 19 include evidence on end results (as contrasted with educational and practice results).

Steps in Evaluation

The abstracts in Volume II illustrate different facets of the four study designs. These facets are related to the identification of a series of steps in carrying out a study of program results. The literature on evaluation is replete with steps considered essential in identifying program results.

Knowles (1970) conceives of four steps in evaluation, to: (a) formulate the questions to be answered, (b) collect the data that will help to answer the questions, (c) analyze the data and interpret them in relation to the questions asked, and (d) propose modifications of the plans, operations, and programs in light of the findings.

The United States General Accounting Office's Assessing Social Program Impact Evaluation: A Checklist Approach (1978), provides a

systematic framework for organizing evidence on program results. It highlights the procedures of: (1) evaluation planning, (2) data collection, (3) data analysis, (4) reporting findings and (5) data disclosure. It places emphasis on the political nature of evaluation, and reminds us that evaluation is not just for program development and improvement but may also be for policy justification or change.

Beder (1979) formulates a more elaborate planning and implementation strategy of eight steps to successful evaluations, as follows:

1. Decide on the purpose and use of the evaluation.
2. Determine what will be evaluated.
3. Acquire and allocate evaluation resources.
4. Establish a proper climate (participation and cooperation).
5. Choose an evaluation design, or approach.
6. Conduct the evaluation.
7. Report the evaluation.
8. Act on the evaluation.

Cunningham (1980) identifies six stages of evaluation, beginning with establishing a steering or advisory committee (a stage consistent with the modis operandi of Cooperative Extension). These stages are:

1. Establish a Steering (or Advisory) Committee
2. Determine Strategy
   --Purpose and Use
   --What to Evaluate
   --Resources and Budget
3. Agree on Design/Approach

4. Conduct/Implement

5. Report Results

6. Act on Results

A steering (planning) or advisory (consultative) committee can be valuable in providing advice and also "consent." A consensus of opinion on a moot point may prove of considerable value when justifying certain plans and actions.

Kappa Systems, Inc. (1979) completed a three-volume review, appraisal and summarization of studies of Extension program effectiveness, including guidelines for improving evaluations of Extension programs. Limitations in study methodology and reporting common to many of the studies appraised were identified and used in developing ten guidelines for future studies. These guidelines, closely related to steps in evaluation, are listed below:

1. Clearly state study purposes.
2. Specify study limitations and/or degree of generalizability.
3. Describe the Extension program being assessed.
4. Relate study questions and measures to program objectives.
5. Discuss the reliability and validity of the measures selected.
6. Establish a link between client outcomes and Extension program delivery.
7. Provide adequate labelling of tables, charts, and graphs.
8. Separate presentation of findings from conclusions.
9. Provide adequate support for conclusions and a comparison if program success or failure is concluded.
10. Balance completeness of report with succinctness of presentation.

The focus of this publication is on Beder's step number five above, and on Cunningham's stage number three. The guideline in Kappa System's list which relates most closely to formulating study design is that of No. 6: "Establish a link between client outcomes and Extension program delivery." A central question in the selection of a study design is precisely this: What logic and facts will be adduced to determine if and to what extent there is a link between client behavior or status and the delivery of an Extension program?

Kappa Systems also distinguishes between study findings on program results (guideline 8) and conclusions about program success or failure (guideline 9). Conclusions on program success require well-formulated criteria, generally based on program objectives on other expectations for the program.

Procedures for formulating and using criteria in evaluating the success or failure of Extension programs (Steele, 1970) are beyond the intended scope of this resource publication. Hence its title refers to studying program results rather than evaluating program results. However, systematic evaluation of program results is usually impossible without sound evidence on the nature and extent of such results. Ascertained program results should lead to answering evaluative questions such as, "Did the results justify the amount of resources invested?" "How badly were these results needed?" "Are other programs which receive similar amounts of resources accomplishing more?" "Should program staff try to increase the magnitude of intended results of this program in the future?"

In summary, our publication on designing studies focuses on only one of the steps or stages in planning, conducting and utilizing

evaluations. As necessary, readers should consult other references which deal with the other steps in evaluation. The intelligent selection or construction of study design depends upon adequate completion of the steps preceding it. Moreover, using study findings intelligently (in evaluation, accountability, and program management and policy decisions) depends upon understanding the nature and limitations of the study design which produced the findings.

Standards for Evaluations

Two major sets of standards for evaluations have appeared recently:

a) The Joint Committee on Standards for Educational Evaluation. Standards for Evaluations of Educational Programs, Projects, and Materials. (1981).

b) Evaluation Research Society Standards Committee. "Evaluation Research Society Standards for Program Evaluation." (1982).

The Joint Committee's Standards are organized under four main headings--utility standards, feasibility standards, propriety standards and accuracy standards--and among the 32 standards included, some 15 refer to designing evaluation studies. One of the strengths of the Joint Committee's standards is not equating high quality program evaluation with technically accurate study methodology. For example, how an evaluation study may be used can be more important in appraising its overall quality than its technical accuracy. Technical accuracy is necessary but does not insure appropriate utility.

Of particular interest is the section of the Evaluation Research Society's standards on structure and design. This section includes the statements:

> --The design for any evaluation cannot be conceived in a vacuum. It is necessarily influenced by logistical, ethical, political, and fiscal concerns and therefore must take these as well as methodological requirements into account.

--For all types of evaluations, a clear approach or design should be specified and justified as appropriate to the types of conclusions and inferences to be drawn.

--For impact studies, the central evaluation design problem of estimating the effects of nontreatment (absence of program) and the choice of particular method for accomplishing this should be fully described and justified.

Prior to the publication of the two above sets of standards, Kappa Systems, Inc. (1979), under contract to Extension Service, USDA, established standards regarding technical accuracy; i.e., regarding substantiation of findings and conclusions in studies of the effectiveness of Extension programs (Appendix B). The 42 studies of Extension program results which are abstracted and analyzed relative to facets of study design all meet the standards which Kappa Systems adopted in consultation with the Science Management Corporation (1979).

Thus, while the accuracy standards employed in this resource publication are only a subset of the standards relevant to appraising studies of Extension program results, they are a fundamental subset. We assume that a study of program results can not have real utility unless it is sufficiently accurate. Our publication does touch upon considerations of propriety and utility, and emphasizes the criterion of feasibility in the selection of study design.

Conclusion

With this brief review of uses, users, steps and standards in evaluation, let us move to an examination of study designs. As stated, the several alternative designs presented are not new to the literature on evaluation; thus our purpose is to bring these designs under close focus, systematically identify design facets and illustrate their value in formulating studies of Extension program results.

References

Beder, H. (1979). Program evaluation and follow up. In P.D. Langerman & D. H. Smith (Eds.), Managing adult and continuing education programs and staff. (pp. 263-283). Washington, DC: National Association of Public Continuing and Adult Education.

Bennett, C. F. (1979). Analyzing impacts of Extension programs. Washington, DC: U.S.D.A. Science and Education (No. ESC-575). (rev. ed).

Cunningham, S. (1981). Monitoring program review. Unpublished paper, Cornell University, Cooperative Extension Service, Ithaca, NY.

Dave, H.R. (1980). A built-in system of evaluation for reform projects and programmes in education. International Review of Education, XXVI, 479-80.

Evaluation Research Society (ERS). (1982). Evaluation research society standards for program evaluation. In P.H. Rossi, (Ed.). Standards for evaluation practices, (pp. 7-19). San Francisco: Jossey-Bass.

Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluations of educational programs, projects and materials. N.Y.: McGraw-Hill.

Kappa Systems, Inc. (1979). Guidelines for improving Extension impact studies. Vol. III. Arlington, VA: Author.

Knowles, M. S. (1970). The modern practice of adult education. N.Y.: Association Press.

Rivera, W. M. (1982). Evaluation in adult education: An international perspective. (Theory, innovation and practice in andragogy, no. 4). Tallahassee, FL: International Institute of Andragogy.

Schmidt, R. E., Horst, P., Scanlon, J. W., & Wholey, J. S. (1977). Serving the federal evaluation market: strategic alternatives for managers and evaluators. Washington, DC: Urban Institute.

Science Management Corporation. (1979). A review and appraisal of studies of Extension program effectiveness. Final Report. Washington, DC: Decision Studies Group, A Division of Scientific Management Corporation.

Smith, N. L. (Ed.). (1981). New techniques for evaluation: new perspectives in Evaluation, Vol. 2. Beverly Hills: Sage.

Steele, S. (1970). Program evaluation - A broader definition. Journal of Extension, 8(2), 5-17.

Stufflebeam, D. L. (1971). A conceptualization of evaluation. Washington, DC: American Educational Research Association (Training Tape, Series C).

United States General Accounting Office. (1978). Assessing social program impact evaluations: a checklist approach. Washington, DC: GAO (PAD-79-2; Exposure Draft).

PART II -- DESIGNS FOR EXAMINING RESULTS

CHAPTER 3: INTRODUCTION TO STUDY DESIGNS

Overview of Selected Designs

A primary assumption underlying studies of Extension program results is that behavioral and status changes of program clientele can be attributed to their participation in an Extension program. An objective of such studies is to obtain sufficient evidence to determine the extent of these attributable changes, i.e., the program's results. Such program results include changes or reinforcements in: knowledge, attitudes, skills or aspirations (KASA change); individual or collective practices; and individual, social, economic and environmental statuses, qualities or conditions (Bennett, 1979).

In order to demonstrate the extent to which specified behavioral or status changes of participants are attributable to an Extension program, it is important that alternate (rival) explanations of these changes be eliminated or taken into account. The design of a study is the major factor in determining how well it can attribute clientele behavioral or status changes to Extension as compared to non-Extension influences. Study designs vary in their ability to account for alternate explanations.

Study designs generally make or imply some form of comparison either within or between groups. When a within-group design is used, it may or may not be known beforehand just what different peoples' exposure to a program has been or the intensity of their program participation.

No attempt is made to establish two or more groups, each with similar characteristics but with different experiences relative to the program. Thus, we focus analysis on one group, or "within a group."

When a between-group design is used, a set of people who are known beforehand to have been exposed in some way to the program are compared to a group that is comparable to the group participating in the program, but which is not exposed to the program or is exposed to some alternate form of the program. Each set of people is then followed by the study through time. Thus, we analyze between these different sets of people or "between-groups."

Within-group study designs do not follow a comparison or control group over time. They may examine change in program participants' knowledge or behavior over time: in such a case, a comparison is made between participants' characteristics at one period of time (usually before the program begins) and at a later time (perhaps at the end of the program). Another means of within-group comparison is to correlate changes in specified behaviors or statuses of program participants (such as change in economic status) with how often they participate. Generally, within-group designs provide only limited design control over rival explanations but instead tend to rely on methods of statistical control. Within-group designs are generally easier to implement than between-group designs because they involve less complicated connections with the program.

Stronger evidence that clientele behavioral or status changes are, in fact, because of participating in a program (as opposed to rival explanations) may be provided when comparisons are made between two or more groups (Campbell and Stanley, 1966). These between-group designs

compare specified characteristics of program participants with those of a comparison or control group with 1) no participation or a varied degree of participation; or, 2) participation in some alternate program. Such comparisons can help assess the role of the Extension program in influencing participants apart from non-Extension influences. If the program group and a comparison group are subject to the same nonprogram influences, any differences in relevant changes between the two groups can be attributed to the program. Two difficulties exist in using between-group study designs: (a) finding a non-program group comparable to the program group, and (b) dealing with the general expectation that all members of the target program audience will have the same opportunity to participate in the program (MAGI, 1979).

The following description reviews four selected study designs commonly used in evaluations of Extension program results. A simple means of describing these study designs is by way of the matrix in Exhibit 3.1. The horizontal labels of this matrix refer to the type of comparison planned--within a group or between groups. The vertical labels refer to whether the design requires pretest or benchmark data.

Exhibit 3.1.
Matrix of Quantitative Study Designs

DESIGN REQUIRES PRETEST OR BENCHMARK DATA	WITHIN-GROUP DESIGNS	BETWEEN-GROUP DESIGNS
No	Survey*	Field Experiment
Yes	Time-Series	Comparison Group

* Survey, is used in the same context as in Bennett's (1979) description of study designs. It is not to be confused with the use of the word "survey" to mean simply a means of collecting data. The term is used both ways in the literature, but for our purposes it refers to a "one-shot" design for studying participants (and sometimes others) only after the fact of program participation.

The two within-group designs included in this publication are the survey and the time-series. More formal names for these designs are, respectively, the "single group posttest only" design, and the "single group pretest/posttest" design (French, 1983).

The two "between-group" designs most commonly used to evaluate Extension programs are the comparison group and the field experiment. More formal names for these designs are, respectively, "quasi-experimental pretest/posttest" design and "true experimental" design (French, 1983).

The Survey Design

When it is decided to ascertain results of an on-going Extension program (one already being implemented) and there are no benchmark data available, it is generally necessary to implement a survey design. A survey design may complete a study of a program's results more quickly than the program that produced these results. This is because, in the survey design, data from clientele are collected only once; e.g., following program participation or at the conclusion of the program. There are two major types of surveys used in the evaluation of program results: perceptual and cross-sectional.

Perceptual Surveys

Perceptual surveys generally show program participants' perceptions of a program's results. Such perceptions are based on participants' reflections or retrospections regarding their behavior or statuses before the program, and their estimates of the change in these behaviors or statuses because of their participation. This perceived "before-to-after" evidence is one way deal with the question of attribution, i.e., what influences created specific behaviors or statuses of participants?

Surveys may also obtain perceptions of program results from key observers of a program who do not participate. Observers provide their perceptions, based on retrospection, of the results of a program for participants and other people to whom the participants relate.

Some study users question the validity of findings on program results based upon perceptions of clientele as to Extension's influence on their behavior or statuses. House (1980) distinguishes between

"subjectivist" and "objectivist" positions or assumptions about the nature, methods and limits of program evaluation. Objectivists consider that evaluation information should be "scientifically objective," achieved through reproducible, quantitative techniques be verified by logical inspection. Those who maintain an objectivist point of view would argue that reflective or retrospective data does not produce adequate evidence of program results. Objectivists question the validity of retrospective data, asserting that what clientele perceive or believe about the results of a program may be influenced by what they want to believe. Furthermore, retrospective data may be invalidated by memory loss or distortion (Rossi and Wright, 1977).

Those who maintain a "subjectivist" point of view base their claim for validity of perceptual data more on an appeal to experience than to "scientific objectivity." Subjectivists hold that human experience is perception -- that it is necessary to discover what program participation means to clientele, and to allow them to connect events in identifying the cumulative effects of multiple experiences in program participation. It is also advanced that perceptual findings on program results are easier for most study users to understand and use than are quantitative documentation of program results (Forest and Marshall, 1977).

An important use of the survey design is to gather perceptions or opinions about the activities and results of Extension programs. In many instances, representative samples of program participants' and key observers' perceptions of Extension program results may be evidence sufficient to meet identified study users' needs in evaluation and decisionmaking.

Opinions are surveyed in may areas, such as: (a) the extent to which Extension program objectives have been achieved; (b) the extent to which Extension and other actors, agencies, etc., have produced given clientele behaviors; and (c) how satisfied Extension clientele are with the results of the programs.

Cross-sectional Surveys

The cross-sectional type of survey is more consistent with "objectivist" approaches to studying program results. Compared with perceptual surveys, cross-sectional surveys rely less on program participants' perceptions and interpretations, and more upon logically and empirically-based quantitative analysis. The same may be said for the time-series comparison group and field experimental designs.

The cross-sectional survey generally includes participants of a program and nonparticipants as respondents. Such cross-sectional surveys include all those -- or a sample of those -- who are in a demographically or geographically defined population (e.g., all farmers or all adults in a state). Such a demographically or geographically defined population will include both program participants and nonparticipants or, the cross-sectional survey may be confined to only program clientele, making sure to include those with different kinds or degrees of program involvement for comparative purposes.

Cross-sectional surveys tend to obtain data on: (a) respondents' self-reported behavior, qualities, or statuses at the time of the data collection (and, perhaps, at an earlier time); (b) factual, situational data such as self-reported participation in Extension and other programs, age, formal education and family status, etc.; and (c) data based on clientele records.

Two possible variations within the cross-sectional survey design should be mentioned.

(1) Data collection only from previously known program participants. Here, the cross-sectional survey can only compare clientele with different kinds and degrees of program involvement.

(2) Data collection from previously known program participants and a random sample of the general population in order to identify and obtain responses from nonparticipants. This approach is a substitute for sampling a defined population and still gather responses from program participants and nonparticipants. (For other variations in survey design, see Clark, 1976).

To deal with the question of how much an Extension program may have contributed to specified clientele behaviors or statuses, data from cross-sectional surveys are subjected to multi-variate statistical analyses, such as: successive statistical adjustments (Rossi and Freeman, 1982), factor analysis, discriminant analysis, multiple correlation and regression and path analysis (Costner, 1971; Mosteller and Tukey, 1977). Such statistical controls are necessary because of the inevitable lack of equivalence in relevant characteristics among program participants and non-participants which may affect specified behaviors or statuses of clientele. It must be determined to what extent such specified behavior or statuses are actually attributable to the program rather than to nonprogram factors.

Appropriate multi-variate analysis of survey data can help determine the conditions which link program participation to results. For example, program participation may be associated with clientele practice

change only among those clientele having certain personality traits (e.g., innovativeness) or socio-economic statuses. Nonprogram factors can obscure or exaggerate study findings on program results, even with multi-variate analyses.

Voluntary participation in Extension programs is a nonprogram factor which is difficult to control statistically. For example, even with complex multi-variate analysis in a survey, it is difficult to tell whether: (a) program participants who are more in contact with an Extension program therefore use more recommended practices, or, (b) program participants who use more recommended practices have sought more frequent contact with Extension and other sources of similar information, or, (c) a combination of (a) and (b) are actually the case over time in some cyclical fashion.

While multi-variate analysis may provide much statistical control over nonprogram factors which confound study findings regarding program results, findings from such multi-variate analyses still may be misleading. Multi-variate analyses rely upon many assumptions about the nature of quantitative variables in the analyses and the typical relationships among these variables. It is often difficult to meet these assumptions, or to know the extent to which they are met: failing to meet such assumptions can substantially affect study findings. The factor of probable differential motivation by program and non-program individuals may be addressed by modeling the process of self-selection into Extension program participation. However, this technique has so far had limited acceptance (Rossi and Freeman, 1982). Thus, authors of survey studies are advised to take great care in (or avoid) inferring that an Extension program is a causal factor in measured clientele outcomes.

The Time-Series Design

When the decision has been made to study results of an ongoing program, and benchmark data on program participants are available and applicable to the program results to be examined, additional data may be collected in order to make time-series comparisons. Additional data collected on participants during or following the program are compared with the benchmark data on the same individual participants.

Time-series designs are used more frequently in Extension to examine results of a new program (implementation will begin in the future) or an ongoing program with a new set of participants. Here, a pretest is administered before the program begins, and these pretest data on clientele characteristics are compared with posttest data on these same individuals.

Time-series studies which collect data only twice (e.g., before and after program implementation) should be distinguished from time-trend studies which repeatedly measure clientele behavior or statuses as they relate to program objectives. An important technique in time-trend studies is to compare time-trend projections of pre-program data with observations of the participants after program implementation. An indicator of program results is the difference between the observed "after" specified behaviors or statuses and those projected based on past overall time-trend data (Hatry, et al., 1981; Rossi and Freeman, 1982).

In time-series studies, comparisons with nonparticipants are not made, but comparisons may be made between the extent of "before-to-after" change among participants with varying types or degrees of

participation in the program. Time-series designs are not as effective as between-groups designs when indicating how much change consistent with program objectives is due to participation in the program or to nonprogram influences. Reports on time-series studies should acknowledge and address the possible influence of applicable non-program factors (e.g., fluctuation in weather conditions on multi-year changes in crop yields). However, time-series studies may provide valid estimates of program results where other possible sources of change in clientele characteristics can be ruled out logically. Also, if several time-series studies of a given program are carried out in multiple sites, each producing similar study findings, rival explanations become less of a threat. Time-series studies often attempt to establish statistical controls in order to determine if or the extent to which participation in an Extension program leads to specified behaviors or statuses.

Program participants' experience in responding to the pretest may indeed be an important stimulus toward measured change in their behaviors or qualities which are sought by the program. This pretesting experience may serve as a stimulus which operates in addition to the involvement in program activities. Special design techniques may be employed to address the possible influence of the pretest.

The Comparison Group Design

When the decision has been made to study results of a new program or of an ongoing program which will have new sets of participants, the comparison group design may be employed. The comparison group design

attempts to establish a high degree of similarity between program participants and a group of nonparticipants. Pre-program information about those who will be exposed to the program and those who will not allows the construction of similar groups. Such "constructed controls" (Rossi and Freeman, 1982) may be accomplished by matching, as much as possible: (a) characteristics of individual pairs of participants and nonparticipants; (b) statistical distributions of salient characteristics of sets of participants and non-participants, or; (c) intact groups of participants and non-participants.

The comparison group design is limited since matching of program participants and nonparticipants is only partial and not complete. Thus, statistical controls are usually added to a degree of constructed control. In using this design, the constructed control and statistical control are combined in order to attempt to remove or reduce differences in the characteristics of participants and non-participants which might affect the clientele behaviors or statuses which are sought by the program.

Authors who use a comparison group design are generally advised to address explicitly the degree of comparability of the two groups concerning characteristics which might pose an alternate explanation to that of program effect. This means identifying factors in addition to the program which may effect the changes specified, so that at least these factors may be accounted for statistically in assessing Extension's degree of contribution to program objectives. Statistical comparisons of, or controls for, program and comparison groups include characteristics such as age, sex, socio-economic status, and aptitudes.

Statistical techniques such as co-variance analysis or multiple regression can correct partially for such extraneous factors. The program and comparison groups may be compared statistically in terms of before-to-after gain scores that are adjusted to account for initial differences in "before scores." However, it may be difficult to compensate for any differences in motivation which led program participants into involvement with Extension (Alexander, 1965). This self-selection bias may not be a problem in some comparison group studies. Possible effects of the pretest or the posttest scores of the program and comparison groups should also be considered.

The Field Experiment Design

When the decision has been made to study results of a new program or of an ongoing program which will have new sets of participants, the field experiment design may be employed. This design is also known as the "randomized group design" (Smith, 1980). The field experiment is so called because it was invented for agronomic field research rather than for biological laboratories where constructed control predominates.

The field experiment requires making the program available to clientele selected randomly (through chance alone) from some potential audience. The part of the audience receiving no exposure to the program is the "control group," i.e., the group which does not participate in the program or participates in a different program.

In the comparison group design, the set of Extension program participants may "volunteer" to participate in the program, while the set of non-participants chooses not to be in the program. This key

difference in the voluntarism of the two sets -- participants and non-participants -- may pose a serious threat to the validity of study findings on program results. As explained above, those who express the desire to participate in the program may have a greater initial motivation to achieve the program's objectives than those who do not volunteer.

But, in the field experiment, it is possible to randomly assign persons who have already volunteered for program participation to: (a) participate in the Extension program ("treatment"), or (b) serve in a control group. In this case, rival explanations for clientele behavior or status can be more completely accounted for than in the previously discussed designs.

In utilizing random assignment, the program and control groups are not expected to be initially equivalent; i.e., all nonprogram factors are not constant between the program and control groups. Instead, the field experiment design randomizes uncontrolled variables among program and control groups (Fisher, 1935). With random assignment, tests for statistical significance can be employed in order to determine the odds that any greater increase in program group achievement over that of the control group was brought about by the presence of the program rather than by uncontrolled factors or chance.

In one type of field experiment, observations within the program and control groups are made both before and after the program. The types of observations in such a field experiment relative to levels of program objectives and clientele outcomes have been depicted graphically (Bennett, 1979). In a second type of field experiment, observations

within the program and control groups are made only after program implementation.

A practical, human relations problem in conducting a field experiment includes the ethics and logistics of gaining the cooperation of control group members. Although the field experiment may be the most effective design for scientifically sound conclusions regarding program results, it is often difficult to randomly assign participants in an Extension educational setting. Those who desire to participate in Extension programs generally expect equal treatment. Denying participation in what seems to be a desirable program, or offering what seems to be a less than ideal program to certain participants may pose ethical questions. However, it may be replied that until a study of program results has been conducted, it is not known whether the program is in reality of benefit to clientele (Bennett and Leonard, 1970). One possible compromise may be to delay offering the program to the control group until after the experimental group has completed the program. A second approach is to assign volunteers to different variations of program intensity, without having a non-program control group.

Summary Comparison of Designs

While the matrix of Exhibit 3.1 provides a simple way for comparing and contrasting the study designs, certain dimensions of each may exist within the other designs. The survey as a technique may actually be a component of all the other designs (Clark, 1976). For example, the time-series design involves collection of survey data from participants

at two or more points in time. The comparison groups design, in turn, incorporates time-series, in making pre- and post-program comparisons. Finally, the field experiment is, in fact, a type of comparison group design though it differs by use of random assignment. Exhibit 3.2 presents a further means of comparing the designs.

Exhibit 3.2: Dimensions of Four Study Designs

	Dimensions of the Design			
Study Design	Data Collected at Conclusion of Program	Pretest or Benchmark Data Manditory	Comparison Between Groups	Random Assignment
Survey	X			
Time-series	X	X		
Comparison Group	X	X	X	
Field Experiment	X		X	X

References

Alexander, D. (1965). A Critique of evaluation. Journal of Extension 3, 205-212.

Bennett, C. F. (1979). Analyzing impacts of Extension programs. Washington, DC: USDA Science and Education (ESC-575).

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Clark, L. P. (1976). Introduction to surveys and interviews. Croton-on-Hudson, N.Y.: Policy Studies Associates.

Costner, H. L. (Ed.). (1971). Sociological methodology. San Francisco: Jossey-Bass.

Fisher, P. A. (1935). Design of experiments. Edinburgh: Oliver and Boyd.

Forest, L. B., & Marshall, M. C. (1978). Impact of Extension in Shawano County: Methodology. Madison, WI: Program and Staff Development, University of Wisconsin - Extension.

French, A. M. (1983). A Guide to evaluation in the Expanded Food and Nutrition Program. Washington, DC: United States Department of Agriculture.

Hatry, H. P., Winnie, R. E., & Fisk, D. M. (1981): Practical program evaluation for state and local governments. Washington DC: The Urban Institute.

House, E. R. (1980) Evaluating with validity. Beverly Hills: Sage.

MAGI Educational Services, Inc. (1979). Developing effective evaluations for adult education programs: A handbook for administrators and evaluators. Harrison, NY: Author.

Mosteller, F., & Tukey, J. W. (1977). Data analysis and regression. N.Y.: Addison-Wesley.

Rossi, P. H., & Freeman, H. E. (1982). Evaluation: A systematic approach (2nd ed.). Beverly Hills: Sage.

Rossi, P. H., & Wright, S. (1977). Evaluation research. Evaluation Quarterly, 1, 5-52.

Smith, H. W. (1980). Strategies of social research: The methodological imagination (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall.

CHAPTER 4: OPTIONS IN DESIGNING A STUDY

Selecting A Study Design

The study designs discussed above offer varying degrees of scientific evidence of program results regarding the extent to which KASA changes, practices, or end results were effected by Extension rather than other sources. However, selecting a study design is a thoughtful process that cannot be technically or mechanically standardized. There are advantages and disadvantages in each design, and what is gained in one may be lost in another.

While many evaluators regard the field experiment as the ideal design study for measuring program results, Cronbach (1982) defends non-experimental designs for program evaluation. He differentiates between the job of the basic researcher and the evaluator, pointing out that advocates of experimental research design often assume that fixed and limited questions will be answered through successive studies over a long period of time. Study users, on the other hand, often call for the illumination of a whole program and its results within a comparatively brief period of time. Gaining a strong cause-and-effect answer to some of the many questions about a program's results through an experimental design must be weighed against the cost of leaving unanswered many other questions about the program's effectiveness. Thus, Cronbach suggests that selection of design for a study may be based on whether to sacrifice breadth of information, that could be gained through a survey, for the sake of more definitive answers to narrower questions. More-

over, the added power of an experiment to attribute observed clientele changes to a program also raises the risk that the study design will interfere with convenient program implementation.

Compared to other designs, the survey generally requires fewer resources for any given scope of programming evaluated. The survey design may be selected in cases where it would be difficult to justify a study requiring a greater amount of time, money, or other resources which would be perhaps required by the other designs. The survey may be selected when the geographic or programmatic scope of the study is extremely wide, thus requiring considerable resources for even the survey design.

A well-designed survey can provide substantial evidence regarding the extent of a program's results. However, assuming that other aspects of a study are of equal quality, the survey is considered by many design experts to provide weaker evidence than between-group designs on whether specified clientele changes are attributable to the program in question (e.g., Campbell and Stanley, 1966). As mentioned earlier in this chapter, the survey design relies either upon respondents' perceptions of program results, or upon statistical controls to achieve evidence of a program's contribution to specified clientele changes. A well-planned and executed survey can be more valid than a poorly executed between-group design.

Cronbach (1982) asserts that: "For any evaluation many good designs can be proposed, but no perfect ones;" and "designing an evaluation is an art." There is not only a choice of designs, but numerous choices within each design, and those choices will and should vary from situation to situation. While it is the study designer's task to produce the

most useful evidence, he or she is working with a specified amount of resources - financial resources, staff resources and respondents' (e.g. program participants') cooperation. Consideration must be given to these factors and several other facets of the study situation when designing a study of program results. Moreover, the study design must be chosen with a view toward how the study will be used in mind. Thus, the utility, feasibility and propriety of a design must be considered in addition to the design's capacity for providing strength of evidence on program results (Joint Committee on Standards for Educational Evaluation, 1981).

You, the reader (as the organizational manager of an evaluation, an evaluator, or advisor to an evaluator) may select or help select ideas from the designs described and the examples provided in this publication. When you begin to choose or plan a design, remember there is no one best design for all situations. The choice will depend upon the interplay of several variables, including information needs of identified study users, time constraints, resources available, creativity, personal judgements, anticipated criticisms, information available and political climate. There will be no automatically correct or incorrect choice of design. Instead, the choice must be a thoughtful, creative process taking into account the realities of the Extension programming and study situations, and balancing the various trade-offs. Recall too, that while we have presented the study designs as individual models, a particular study may require a combination of designs.

In addition to the following overview of facets of study designs, suggestions on design selection may be found in Chapter 6 of this volume.

Overview of Facets of Study Design

Ten facets of study design which are addressed below may assist you in choosing or formulating a study approach. The facets ennumerate various considerations in choosing and executing a study design. While presented as separate considerations, the facets are interrelated and must be considered in light of each other. We have not attempted to be exhaustive in our coverage of design facets: other facets not treated here are methods of reporting to intended study users, and ethical considerations.

The facets of study design which we address are:

1. Program situation
2. Type of study objective
3. Resources
4. Time frame
5. Scope
6. Data sources
7. Sampling strategies
8. Data collection
9. Nature of data
10. Analysis of data.

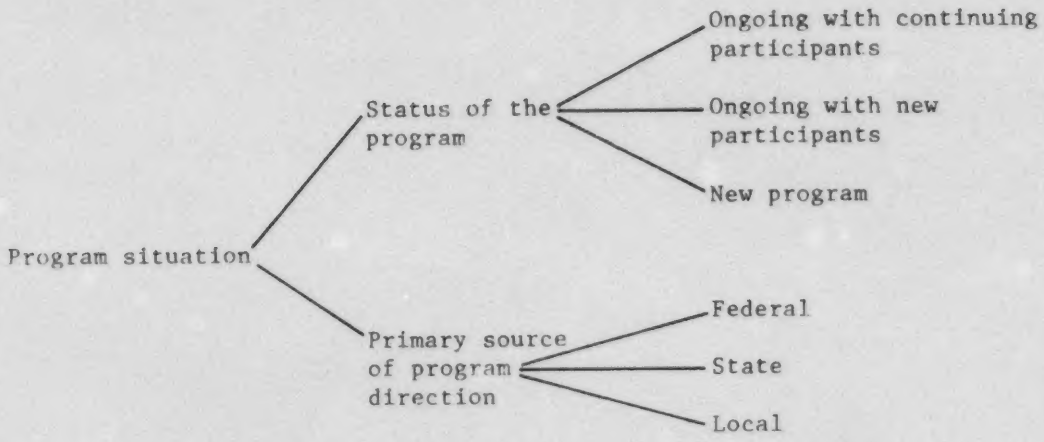
We will discuss the facets above in the following pages with emphasis upon choosing or determining a specific option or options within each facet. Then, in Chapter 5, some of these various options will be illustrated for each of the four study designs by citing previous studies of Extension program results abstracted in Volume II.

We have presented the following section as though readers will be determining, alone or as a group member, which of the options for study methodology to select. It is assumed that program leaders and specialists will be able to identify with this perspective and apply it to their roles in planning studies, e.g., manager for Extension sponsored program evaluation or member of an advising or steering committee for an evaluation.

The key to selecting the appropriate options for designing a study is two fold: (a) ascertaining who needs what information about program results for what purposes, and (b) fitting this "demand" for information with the options for "supply" of such information which are actually feasible. The question, "Why is the study of program results needed?" must be reconciled with the answer to, "What kind of study is feasible?"

The ten facets are presented in separate figures (4.1-4.10) and are analytically broken down into options to help you visualize the choices to be made with regard to each facet; each illustration is then followed by various "considerations" that spell out the visual analysis.

Figure 4.1: The Facet of Program Situation



To determine the situation regarding the program selected for evaluation consider:

1. What is the status of the program?

   Ongoing already with clientele who are continuing to participate?

   Ongoing already, but to be conducted with a new set of participants?

   A new program to be initated with an audience who will begin to participate in the program?

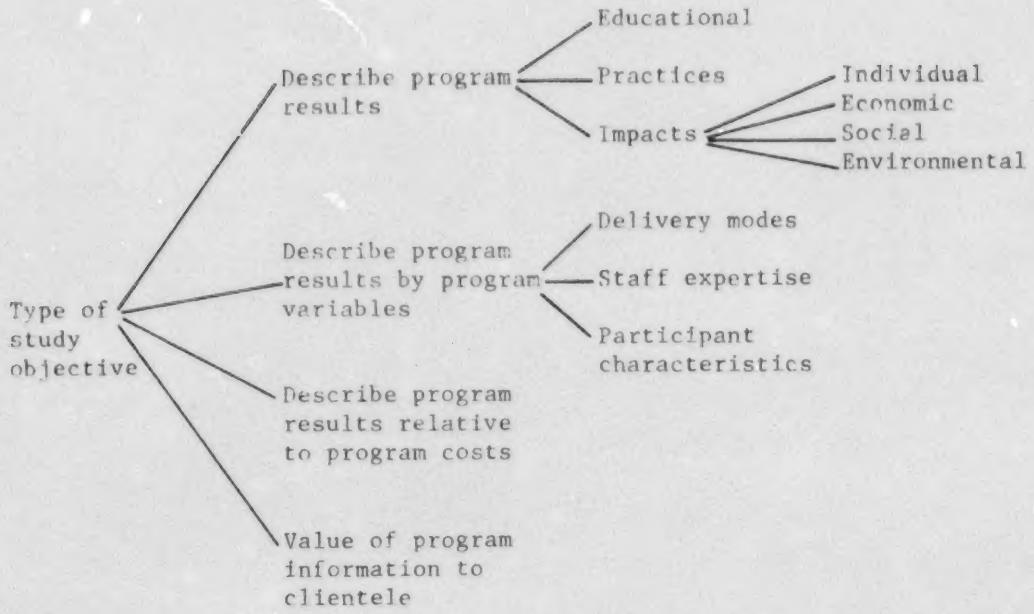
2. What is the primary source of program direction?

   The federal level of Extension (because the program is federally funded)?

   The state level (because the state or campus funds the program)?

   The local level (because of county or city resources supporting the program)?

Figure 4.2: The Facet of Type of Study Objective



To choose a type of study objective, consider whether the intent is to:

1. Describe how the program participants or others changed or benefitted as a result of the program?

   What have they learned?
   What practices have they adopted, changed or continued?
   What are the individual, economic, social and/or environmental consequences?

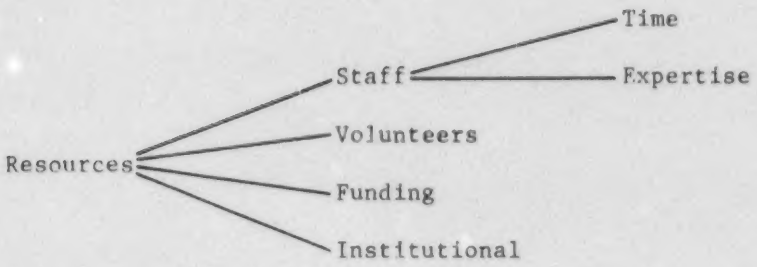
2. Describe the results of a program relative to its variables?

   Is the delivery mode related to results?
   Is degree of staff expertise, or some other attribute, related to results?
   Are characteristics of participants related to program results?

3. Determine the least expensive way to achieve intended program results or determine extent of results that can be obtained for a given cost.

4. Determine the value perceived by clientele of the information supplied by Extension and by other sources for their decision-making.

Figure 4.3: The Facet of Resources for the Study



To determine the resources available consider:

1. Which Extension staff are available to conduct the study?

   How many hours are available from Extension staff?

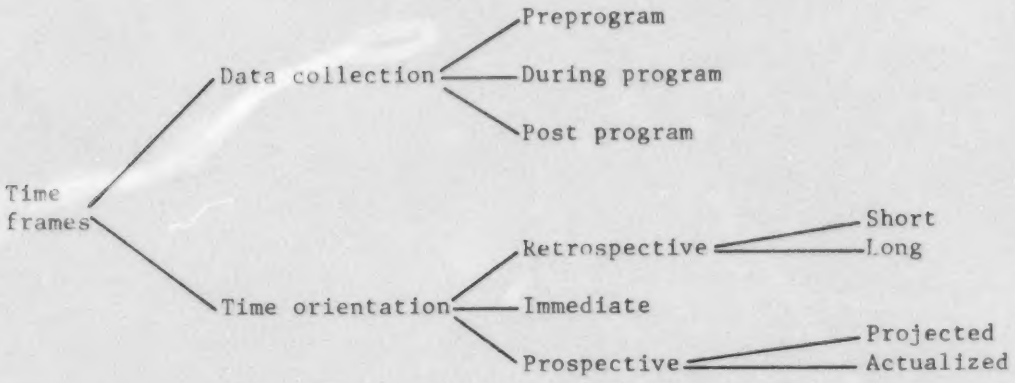
   What types of expertise are available?

2. Is volunteer help available?

3. Will funds be available in order to retain non-Extension staff to conduct or help conduct the study? How much money is available?

4. What types of university, Cooperative Extension Service or academic department support will be provided?

Figure 4.4: The Facet of Time Frames



To determine the time frames for the study, consider:

1. At what point(s) in time will the data be collected?

   Prior to the program?
   During the program?
   After the program?

2. What will be the time orientation(s) relative to program results?

   Examination of past KASA change, practices or end results?

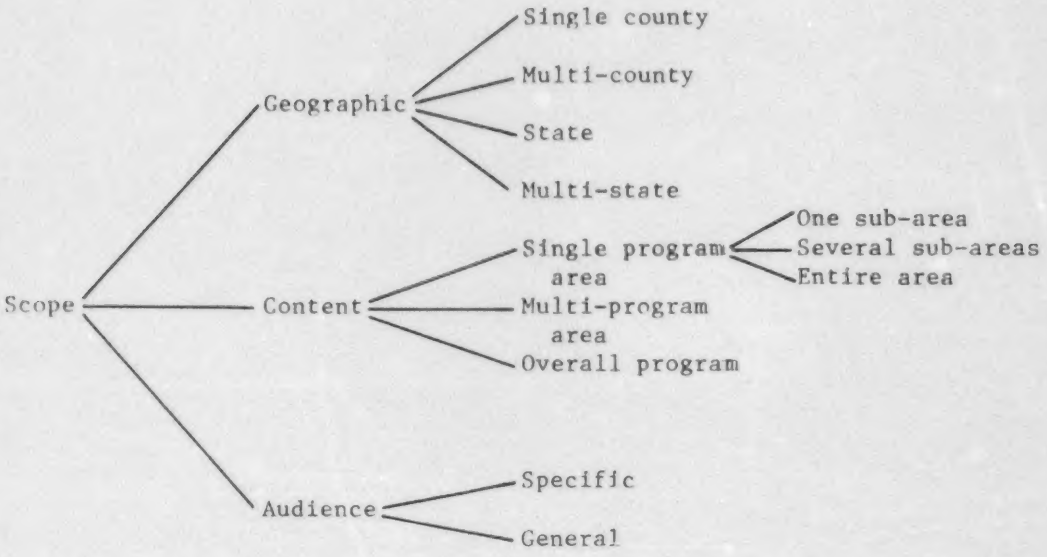
   From a program conducted recently?
   From a program conducted a longer time ago?

   Examination of the immediate results of a current program?

   Examination of longer term results from a current or future program?

   Collection of data now on clientele's intended behavior?
   Collection of data at a later point in time on clientele's actual behavior?

Figure 4.5: The Facet of Study Scope



To determine the scope of the study, consider:

1. The geographic area to be covered by the study?

   A single county or unit within the county?
   More than one county? If so - which ones?
   The entire state?
   More than one state?

2. Which program area(s) will be examined?

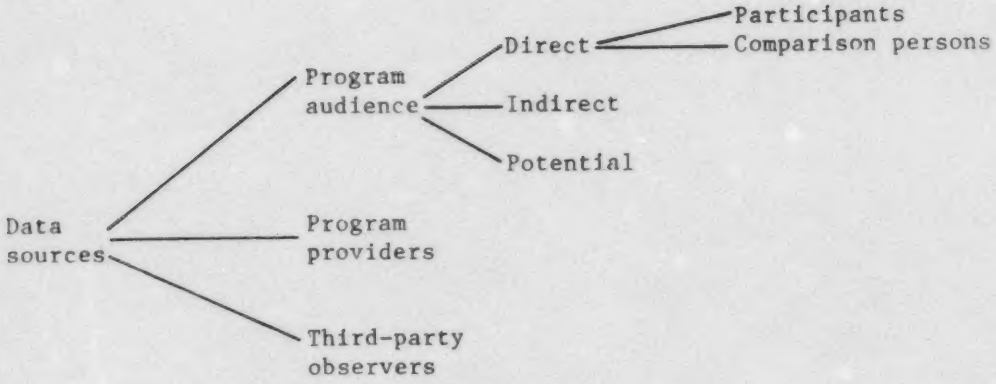
   One program area?

      A single sub-area?
      Several sub-areas?
      The entire program area?

   More than one program area?
   The overall Extension program?

3. Whether the audience population will be the specific target audience(s) of the program or a more general audience which includes the program's audience.

Figure 4.6: The Facet of Data Sources



To determine the data sources for evidence on program results, consider:

1. Which audiences will be asked to supply data?

   Program participants and/or comparison or control group persons?

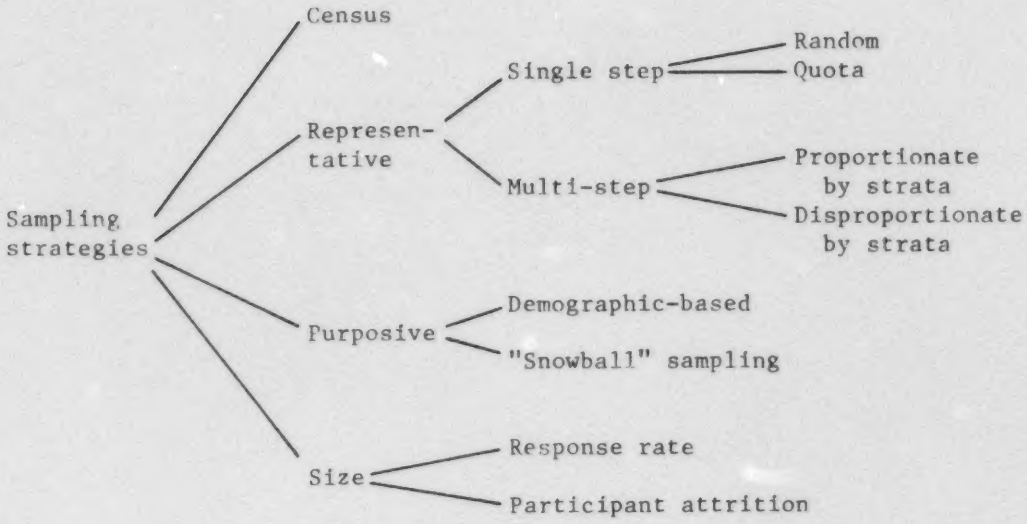
   Persons having indirect contact with the program through the program participants?

   The potential but yet unreached program audience?

2. Will data be sought from the program providers?

3. Will data be gathered from key persons and/or community leaders who observe a program but not necessarily as clientele.

Figure 4.7: The Facet of Sampling Strategies



To choose a strategy for selecting a sample of program providers and/or audience, consider:

1. Whether to study the entire population?

2. Whether to select a representative sample of the population?

   Randomly from the entire population?
   By quota from the entire population?
   Randomly from stratified sections of the population?
      With stratified areas represented proportionately?
      With stratified areas represented disproportionately for the sake of sampling efficiency?

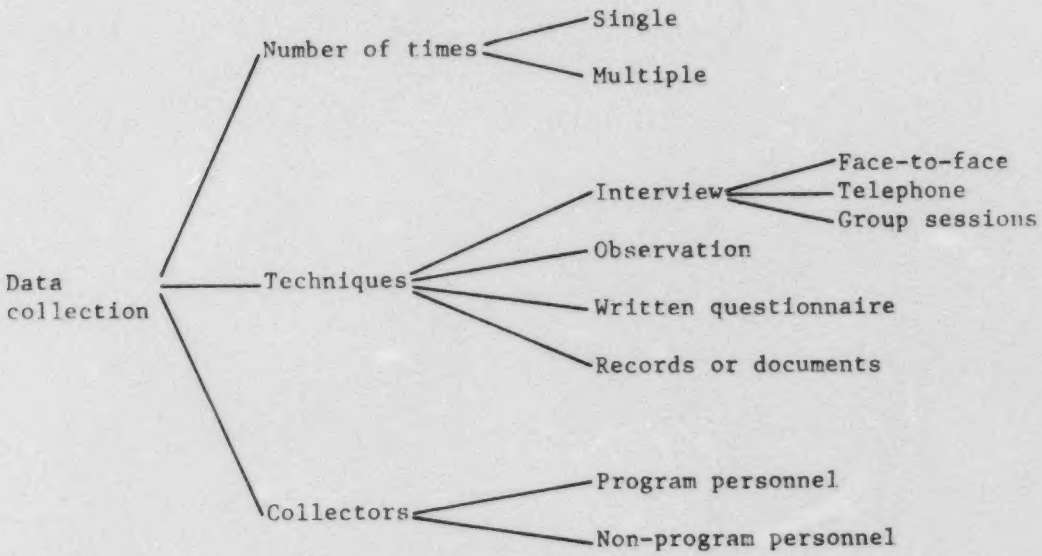
3. Whether to select a purposive sample, where the goal is to gain exploratory knowledge rather than to generalize?

   Demographic-based selection of sampling units, in order to assure examination of varying conditions.
   "Snowball" or other reputationally-based methods of sample selection?

4. How large a sample size is required to permit the precision and analyses desired?

   Considering the expected response rate?
   Considering reductions in numbers of participants by the end of program implementation?

Figure 4.8: The Facet of Data Collection



To determine how to collect the data, consider:

1. How many times to collect data from each information source.

   Only once?

   More than once?

2. What technique will be used to collect data?

   Interview program providers and/or participants?

   In person?
   By phone?
   In group sessions?

   Directly observe behavior within the study sample?

   Distribute a questionnaire?

   Examine personal or public records and documents?

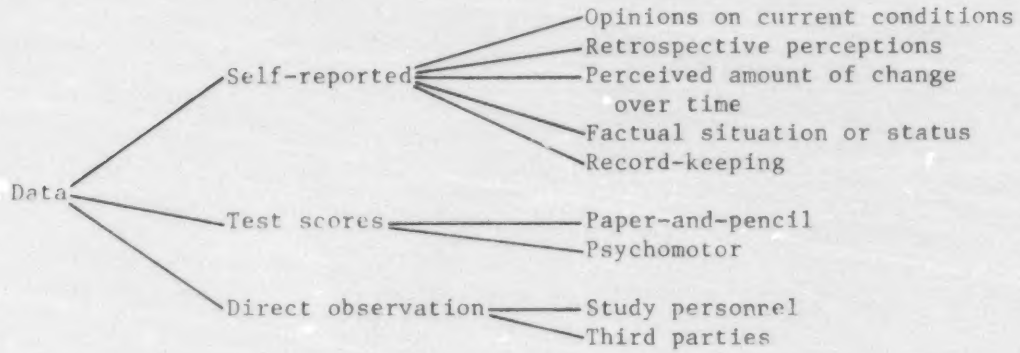
3. Who will collect the data?

   Paid program personnel?

   Volunteer program personnel?

   Nonprogram personnel?

Figure 4.9: The Facet of Nature of Data



To determine the nature of the data to be collected and analyzed, consider:

1. Will the data be self-reported?

   Data on respondents' perceptions or opinions regarding their current behavior, status or surroundings?

   Data on recollections (retrospections) of behavior, status or conditions?

   Data on changes in behavior, status or conditions based on comparing retrospective perceptions and current perceptions.

   Data on factual (objective) or status (e.g., occupation or age).

   Data generated from records based on respondents' observations?

2. Will the program providers and/or audience be tested for knowledge or skills?

   Through paper-and-pencil tests?

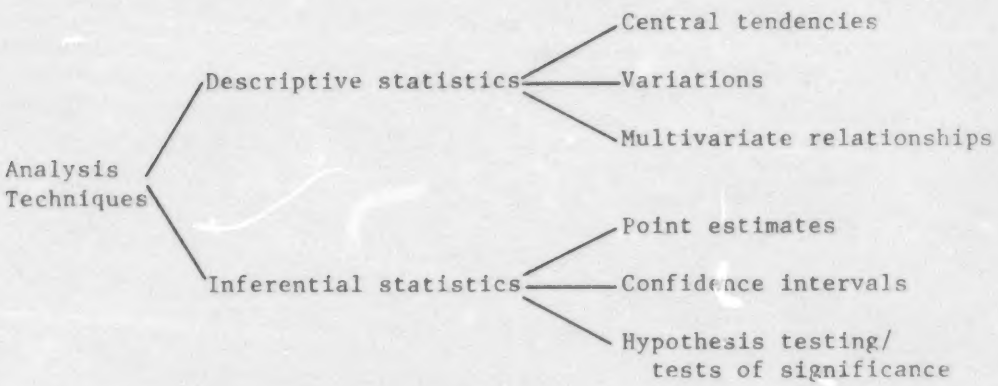
   Through psychomotor skill examinations?

3. Will the data on program providers and/or audience be generated from direct observations of behavior or behavior traces or other data sources?

   Data based on observations by study personnel?

   Data based on third parties' (e.g. family members', community officials) observations or records?

Figure 4.10: The Facet of Analysis of Data



To choose methods for analyzing data to ascertain program results, consider:

1. What types of descriptive statistics will be used?

   Will there be averaging generalizations about clientele outcomes?

   Will variability among participants be examined with regard to clientele outcomes?

   Will associations between clientele outcomes and specific program and/or audience variables be examined?

   Will there be an examination of the simultaneous relation of two or more program and/or audience variables to clientele outcomes, i.e., the use of statistical controls?

2. What types of inferential statistics will be used?

   Are there plans for providing "best estimates" of clientele outcomes among the total audience for the program, based on findings from the study sample?

   Are there plans to provide probability margins for best estimates of clientele outcomes among the total program audience, based on sample findings?

   Are there plans to test hypotheses that stated clientele outcomes have occurred among the total program audience, based on sample findings?

The foregoing facets, and options within these facets, have been discussed by many authors in the fields of educational and social research and public program evaluation including Rossi and Freeman (1982); Smith (1980); and Weiss (1972). Our intent has been to compile and graphically present a type of "checklist" to aid study design formulation within a wholistic picture of the study planning process.

The following chapter will offer a more detailed explanation of each design as it relates to the facets outlined above.

References

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Cronbach, L. J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluation of educational programs; Projects and materials. N.Y.: McGraw Hill.

MAGI Educational Services, Inc. (1979). Developing effective evaluations for adult education programs: A handbook for administrators and evaluators. Harrison, NY: Author.

Rossi, P.H., & Freeman, H. E. (1982). Evaluation: A systematic approach (2nd ed.). Beverly Hills: Sage.

Smith, H. W. (1980). Strategies of social research: The methodological imagination (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Weiss, C. (1972). Evaluation research: methods of assessing program effectiveness. Englewood Cliffs, N.J.: Prentice-Hall.

CHAPTER 5: STUDY DESIGNS, FACETS AND EXAMPLES

In this chapter we analyze each of the four study designs according to facets described in the previous chapter. We cite specific studies abstracted in Volume II in order to exemplify some of the options within the facets. Chapter 5 may serve to familiarize readers with the range of example studies in Volume II.

Survey Designs

Types of Study Objectives

Survey studies describe program results in terms of different levels of evidence. Studies abstracted in Volume II exemplify program participants' opinions as to the extent of their knowledge change (Steele and Everson, '78), attitude change (Forest, '77b), increase in skills (Tait, '69; Rockwell et al., '82), or aspirations (Rockwell et al., '82), as a result of program participation. Perceived practice change also is examined using this design (Goetting, '82; Williams, '78; and others). End results examined by a survey include perceived economic benefits (Glass and Reese, '76), perceived improvements in health and safety (Forest, '77), and increased self-confidence (Rockwell et al., '82; Steele and Everson, '78).

Surveys may describe program results according to such factors as presence or degree of clientele participation in an Extension program (Brown, '67, Dardeau, '77).

Survey studies may also show clientele's perception of the role of Extension in providing specified types of information. This type of survey generally gathers data from a representative sample of a defined population and may be called a "value of information survey" (Awa and Crowder, '77; Bogenschneider, '77).

Time Frames

Surveys lend themselves to the retrospective examination of program results--short term (Kanarek, '78; Tait, '69) and long term (Rockwell et al., '82). Some surveys include prospective data on the actions of clientele, e.g. stated intentions or aspirations to engage in specific practices which are recommended by an Extension program (Goetting, '82).

Scope of Studies

Some of the survey examples included in Volume II examine a program in only one geographic unit or county (Awa and Crowder, '77), others exemplify multi-county surveys (Rockwell et al., '82) and others statewide surveys (Goetting, '82; Steele and Everson, '78).

While the majority of the survey design examples in Volume II focus on a single subject matter area, Rockwell et al. ('82) and Summers and Zeller ('78) examine more than one subject matter area. Forest and Marshall ('77) cross program areas and examine results of the overall Extension program in one Wisconsin county as perceived by participants.

Data Sources

Besides collecting evidence from program participants, surveys may also gather data from third parties, for example, parents of 4-H participants (Steele and Everson, '78; Summers and Zeller, '77). Data

collection from the potential audience for a program is exemplified by Awa and Crowder ('77) and Bogenschneider ('77). Dardeau ('77) and Forest and Marshall ('77) also demonstrate collection of data from a random sample of a geographically defined population, to discover both the total extent of participation in a program and the self-reported results of participation in that program.

Data may be collected from community leaders as observers of program results (Forest et al., '76). Data may be obtained from public records as a supplement to survey data from clientele (Frye and Miller, '76).

Sampling Strategies

Studying a small number of participants, Forest ('77b) and Glass and Reese ('76) included the entire population in surveys. Simple random samples were used by Bogenschneider ('77) and Kanarek ('78). Stratified sampling was used by White and Ladewig ('79) and Tait ('69). Williams ('78) selected sample size expressly to be large enough to insure a sufficiently accurate representation of the population.

Data Collection

Techniques for data collection in the survey include the face-to-face personal interview (Dardeau, '77), the telephone interview (White and Ladewig, '79), mail questionnaires (Glass and Reese, '76), or some combination of the above (Steele and Everson, '78). Brown ('67) demonstrates how survey data may be supplemented with data from direct observations made by interviewers. Forest and Marshall ('77) demonstrate the use of non-program personnel as data collectors.

Nature of Data

Self-reported opinions on respondents' current farm and woodland production practices were collected by Brown ('67) and Dardeau ('71). Most of the survey examples in Volume II are based on self-reports of program participants regarding the amount of perceived change in their behavior or change in qualities due to their Extension program participation (e.g., Summers and Zeller, '77). White and Ladewig ('79) elicitated respondents' (a) retrospective perceptions regarding their past behavior, and (b) perceptions of their current behavior, comparing self-reported behavior before and after participation statistically.

Analysis of Data

Most of the survey examples in Volume II utilize percentage distributions of self-reported program results for participants as a whole. However, some of the examples compare the distribution of responses of program participants exposed to different program delivery modes (White and Ladewig, '79). Length of program participation is correlated with the extent of self-perceived program results (Rockwell et al., '82). Dardeau ('77) examines associations between participants' characteristics and adoption of specified practices.

Summary Statement on The Survey Design

The survey is a valuable design for conducting studies of perceived program results. If nonparticipants are included in a survey, a variety of multi-variate statistical analyses can be employed to learn Extension's contribution to clientele behavior or status. Surveys can be comparatively simple, inexpensive and less time consuming to implement

than the other designs described in this resource. Also, the survey is flexible and can be used to secure evidence of results of a program that is already in progress.

Volume II contains abstracts of 17 studies which utilized a survey design. (Six of the 17 include findings on end results, i.e. program impacts.) Five additional survey studies not abstracted in Volume II are referenced in the section on Further Readings-Survey, which appears at the end of this chapter.

Time-Series Designs

Types of Study Objectives

Change in clientele knowledge and attitudes has been measured using pre- and posttesting (Crowe et al., '76). Kingdon and Toensmeyer ('75) compared knowledge change as well as practice change at six-month intervals using self-reports of food consumption to determine a program's point of diminishing returns. Time-series designs have documented program end results in terms of economic gain (Strickland et al., '76), increased agricultural production (Ladewig and Edmondson, '72) and positive self-concept (Marks, '71).

Time-series designs have also been employed to examine program results as related to various delivery modes (Texas Agricultural Extension Service ,'74), program intensity (Carter, '80) and degrees of staff expertise (Marks, '71). Marks ('71) also examined program results in relation to specified audience characteristics.

Time Frames

In addition to collecting data both before and after program participation (Hartman and Brown, '70), some time-series studies collect data for a third time (perhaps months after the close of participation), in order to check for maintenance of any clientele progress (Marks, '71). Other time-series studies collect data at intervals throughout program participation (Kingdon and Toensmeyer, '75). Some time-series studies obtain data on program delivery or program participation over the duration of a time period of program implementation (Carter, 1980).

Scope of Studies

Time-series studies can be carried out within a single geographic unit (Crowe et al., '75) or in multi-units (Ladewig and Edmondson, '72). The time-series design is particularly applicable when studying a single subject over a wide geographic area (Carter, '80), since limiting the program subject matter helps to simplify the complexities involved in multiple data collections.

Data Sources

In addition to collecting data directly from the program participants, supplementary perceptual data and opinions may be collected through surveys of community observers (Crowe et al., '75; Balliet, '78). Supplementary survey data may also be collected from those who are in turn "clientele" of the program participants (Brown and Nelson, '77).

Sampling Strategies

Kingdom and Toensmeyer ('75) sampled families from each of three counties proportionately to the number of program families in these respective counties, in a proportional random sample. Carter ('80) employed a statewide random sample of dairymen in a survey combined with a time-series study.

Data Collection

Pre- and post-questionnaires as employed by Crowe et al. ('75) are typical of a data collection technique commonly used in time-series studies. Records kept by clientele over a period of time may be employed to study economic impacts (Balliet '78) and changes in agricultural production (Ladewig and Edmondson, '72). Pre- and post-program data gathered in personal interviews can provide evidence on practice change. Ladewig and Edmondson ('72) collected post-program only data on participants' perceptions of change in their status in order to supplement a time-trend analysis. Likewise, a mail questionnaire sent out by the Texas Agricultural Extension Service ('74) provided auxiliary survey data to the time-series analysis of the study.

If data on Extension program staff or clientele can be obtained from an existing database rather than through a special data collection effort, the task of data collection may be simplified. For example, Carter ('80) exploited state-wide data on Extension program activities, by county units, collected routinely during a period of time through the Tennessee Extension management information system.

Analysis of Data

The majority of the time-series studies presented in Volume II express program results as percentage changes. Marks ('71) used analysis of variance to test whether scores indicating self-esteem increased significantly (statistically) from the pretest to the posttests, and whether there was a difference in program results by professional and non-professional Extension Staff.

In determining the relationship between program results and other variables, Hartman and Brown ('70) used regression analysis to determine whether proximity of Extension demonstration farms predicts change in farmers' practices. Multiple regression analysis employed by Crowe et al. ('75) determined the relationship between program results and frequency of participants' attendance at Extension meetings. Carter ('80) investigated the percent of variance in dairymen's acceptance of recommended dairy production practices that could be explained by extent of contact with Extension through applying linear and curvilinear regression analyses.

Summary Statement on the Time-Series Design

The time-series design may be selected to obtain firmer measurements concerning changes in behavior, qualities or status of program participants than data based on perceived changes. Compared with the survey, a longer time period is generally required for defining program results. The design may be suitable for examining program results when using a comparison group is impractical or unnecessary.

Ten abstracted studies included in Volume II use the time-series design for assessing program results. (Six of these 10 studies include

findings on program end results or impacts.) The reader may also wish to refer to four time-series studies which are not abstracted in Volume II but are referenced in the Further Readings at the close of this chapter.

Comparison Group Designs

Types of Study Objectives

The comparison group design has been used to compare program participants and non-participants with respect to change in knowledge (Subaima, '61), attitudes (Street, '73), and aspirations (Green et al., '72). Practice change has been similarly examined (Aldrich, '75). End results measured in comparison group designs have included "quality of life" (Street '73), labor efficiency (Alexander and Longest, '62), and income (Aldrich, '75). Comparisons of program results have been made regarding participants who have received: different program activities, i.e., treatments (Mauk, '75); program delivery by staff with varying levels of expertise (Street, '73); or, instructional activities in different learning environments (Boone and White, '76).

Time Frames

Posttesting in comparison group studies can occur immediately following a short term program (Boone and White, '76), a longer-term program (Aldrich, '75), at set intervals throughout the program (Green et al., '72) or at some interval of time following program participation (Subaima, '61; Marks, '71).

Scope of Studies

Comparison group designs have been used to examine the results of programs in single counties or geographic units (Gruenhagen, '69) and in multi-county studies (Alexander and Longest, '62). While most comparison group studies focus on one subject within a program area, Aldrich ('75) used the design to examine the effects of an integrated agricultural and home economics program.

Data Sources

Data collected directly from program participants and non-participants may be supplemented with observations by third party observers (Subaima, '61). Most comparison groups represented among the studies were based on their objective similarity to the respective program groups of the studies (e.g., Alexander and Longest, '62; Mauk '75; Street, '73). In contrast, Green et al. ('72) established comparison groups comprised of designated friends of members of the program groups.

Sampling Strategies

A census of participants and comparison individuals is useful when the study concentrates on a program of limited size and/or small geographic area (Mauk, '75). Random sampling within stratified sections of a general population was employed by Gruenhagen ('69), while Alexander and Longest ('62) used a modified random sample of program participants and non-participants. Street ('73) employed stages of sampling with a convenience sample followed by a stratified sampling procedure.

Data Collection

Techniques for collecting the data can include: the personal interview (Aldrich, '75); the telephone interview; the mail questionnaire (Gruenhagen, '69); records and documents (Alexander and Longest, '62); and testing (Boone and White, '76). While program personnel may assume the responsibility for collecting the data, Gruenhagen ('69) and Green et al. ('72) employed data collectors unrelated to the program.

Nature of the Data

The self-reported data of Alexander's and Longest's ('62) study was supported by records of farm income, while Subaima ('61) supplemented test data on youth with teachers' perceptions of these youth. Self-reported data may be global in nature (Street, '73) or specific to a particular learning or practice (Gruenhagen, '69).

Analysis of Data

Because program and comparison groups often differ on pretest scores, the groups may be compared statistically in terms of gain scores adjusted to account for initial differences (Boone and White, '76). Analysis of covariance with statistical tests may determine the significance of the differences between or among the groups (Gruenhagen, '69; Subaima, '61; Alexander and Longest, '62).

Summary Statement on the Comparison Group Design

The comparison group design can help to explain the possible influence on clientele outcomes of factors other than the Extension

program. However, there may be problems in gaining the cooperation of nonparticipants and in achieving sufficient actual or analytic comparability between program and comparison groups. Such problems make the comparison group design more difficult to implement than the two within-group designs. Insufficient comparability may lead to erroneous conclusions regarding program results.

Eight studies are abstracted in Volume II which employed a comparison group design. (Six of the eight studies include findings on program impacts or end results.) An additional six studies are cited in the Further Readings.

Field Experiment Designs

Types of Study Objectives

The field experiment is effective in demonstrating Extension program-induced KASA and practice change, as all of the Volume II abstracts illustrating the field experiment design show. Bowering et al. ('78) studied end results of a nutrition program with a field experimental design. Comparisons of program results in relation to type of delivery mode were made by Honnold et al. ('81). Trent and Kinlaw ('76) also used the field experiment to compare delivery mode effectiveness as did Bowering et al. ('78). Honnold et al. ('81) uniquely compared program results to Extension cost per program participant. Program results can also be examined in relation to participant characteristics (Trent and Glass, '77; Scherer, '77).

Time Frames

It is possible in one variation of the field experiment to collect data regarding program participants only at the conclusion of the program. However, pretest scores, or benchmark data, can facilitate field experimental analyses by providing evidence on degree of initial equivalence of the participant and non-participant groups (Trent and Glass, '77).

Sometimes pretesting may not be desirable due to its "reactive effect." That is, studies of pretested individuals indicate that they tend to improve their performance on the posttest regardless of their type of program exposure, or even with no exposure to the program (Trent and Kinlaw, '76). The pretest itself may actually conceal the program (treatment) effect by sensitizing members of the control group so that their posttest scores are affected. Therefore, the pretest is often not administered when conducting a field experiment (Scherer, '77).

Scope of Studies

The field experiment design can be applied to studies of any geographic scope. Bowering et al. ('78) studied clientele of a single baby clinic, while other field experiments have examined program results in several counties (Trent and Glass, '77), in sample counties of the entire state (Trent and Kinlaw, '76), and across two states (Honnold et al., '81).

Sampling Strategies

Field experiments that are narrow in geographical scope may study the entire program population (Bowering et al., '78), while those that

are broad in geographical scope with a large number of participants may use simple random sampling (Scherer, '77) or a two-step random sampling process (Trent and Glass, '77). On the other hand, Honnold et al. ('81) conducted a census of participants even though the program had over 3500 participants enrolled.

Data Collection

As demonstrated in five of the field experimental studies in Volume II, the personal interview can lend itself to this design. Paper and pencil testing is useful (Trent and Glass, '77) as well as a mailed questionnaire (Honnold et al., '81).

Nature of the Data

The field experiments abstracted in Volume II illustrate the use of tests to measure knowledge change (Honnold et al., '81), scales to measure attitude change (Trent and Glass, 1977) and structured self-reporting to measure practice change (Yerka, '74). An exception to the tendency for field experiments to use highly quantified data is the examination of participants' and nonparticipant controls' perceived practice change (Warren et al., '66).

Analysis of Data

As shown in all of the Volume II abstracts, some form of statistical analysis compares differences in the mean test scores of the program and the control groups. This analysis determines if these differences are statistically significant. Several studies also examine the

correlation between variations in participant behavior and characteristics: (a) of the Extension program providers (Yerka, '74), and; (b) of the program participants (Trent and Glass, '77; Scherer, '77).

Summary Statement on the Field Experiment Design

The field experiment can provide strong evidence of an Extension program's results by accounting for other factors which may have affected clientele. However, the field experiment is more effective in studying structured, narrower-scope programs than programs with a wider scope in content or subject matter and with less structure. Use of the field experiment requires close cooperation between program managers and program evaluators, in order to satisfy study design requirements. Where it is feasible and ethical to use random assignment, field experimental findings can be very scientifically persuasive in measuring Extension program results.

Volume II contains abstracts of seven studies using a field experiment design. (Only one of the seven studies includes findings on program impacts or end results.) An additional study is referenced in the Further Readings.

General Comments and a Reminder

The citations above, to studies in Volume II which exemplify options within the design facets, are not at all exhaustive: many more options within the facets could be exemplified by the studies. We hope that a sufficient number of facet options have been explored to suggest the richness of the Volume II abstracts as a resource for planning future studies of Extension program results.

Also, it should be noted that the studies contain other interesting and helpful aspects for planning studies. For example, relative to users and uses of studies, Forest's ('77b) technique of identifying study users' criteria for judging success of the program is noteworthy; and, the concerns by legislators and administrators which led to Goetting's ('82) study are instructive also regarding users and uses. Forest and Marshall ('77) also identify a study user's (administrator's) specific request as the basis for their study of an entire county Extension program.

Before you inspect more closely the abstracts in Volume II, we wish to remind you that these examples were selected from among those available through the sources previously discussed. They are not necessarily intended to serve as "model" studies, but rather as examples of studies using one of the four study designs presented. The examples show how studies of Extension program results using these four designs actually work.

Note also that several of the studies, while grouped within a particular design, really used more than one design. And, some of the studies which are grouped within a particular Extension program area actually encompass program activities in more than one program area.

Further Readings - Survey

Bennett, C.F. (1982). Reflective appraisals of programs (RAP): An approach to studying clientele-perceived results of Cooperative Extension programs. Ithaca, NY: Media Services, Cornell University.

Dillman, D. (1978). Mail and telephone surveys. N.Y.: John Wiley and Sons.

Erickson, S. & Olien, C. (1961). Evaluation of TV clothing series "What shall I wear?" Abstracted in Science Management Corporation. (1979). Final report: A review and appraisal of studies of Extension program effectiveness. Vol. II. (pp. 155-157). Arlington, VA: Author.

House, E.R. (1980). Evaluating with validity. Beverly Hills, CA: Sage.

Issacs, N.W. (1978). Attitude and value study of the 4-H market lamb project in Gallatin County. Abstracted in Kappa Systems, Inc. (1979). Extension program impact findings from selected studies, conducted from 1961 to 1978 Appendix to Vol. II, Supporting Abstracts. (p. 117-118). Arlington, VA: Author.

Katz, D., Gutek, B.A., Kahn, R.L., & Barton, E. (1975). Bureaucratic encounters: A pilot study in the evaluation of government services. Ann Arbor, MI: University of Michigan, Institute for Social Research.

Lee, R.L. (1967) The flow of information to disadvantaged farmers. Doctoral dissertation, University of Iowa. Abstracted in Science Management Corporation. (1979). (as referenced in Erickson above). (pp. 16-18).

Payne, S.L. (1951). The art of asking questions. Princeton, NJ: University Press.

Perry, S.F. (1976). Factors affecting degree of success attained by families in a pilot educational housing program in Pittsylvania County, Virginia. Abstracted in Kappa Systems, Inc. (1979). (as referenced in Issacs above). (pp. 57-59).

Wing, K.E. & Melniker, N.E. (1977). Evaluation of Maine's improving rural homes project. Abstracted in Kappa Systems, Inc. (1979). (as referenced in Issacs above). (pp. 111-113).

Further Readings - Time-Series

Abbott, R., Adler, K., & David, C. (1978). An evaluation of the Detroit urban gardening program. Abstracted in Kappa Systems, Inc. (1979). (pp. 9-12). Findings from selected Extension program impact studies conducted from 1961 to 1978. Appendix to Vol. II, Supporting Abstracts. Arlington, VA: Author.

Hardee, J.G. (1963). Evaluation of an educational program with part-time farm families, Transylvania County, North Carolina 1955- 1960. Abstracted in Science Management Corporation (1979). Final Report: A Review and Appraisal of Studies of Extension Vol. II. (pp. 9-12). Washington, DC: Author.

Hatry, H.P., Winnie, R.E., & Fisk, D.M. (1981). Practical program evaluation for state and local government. Washington, DC: Urban Institute.

Howard, G. (1980). Response shift bias: A problem in evaluating interventions with pre-post self-reports. Evaluation Review, 4, 93-106.

Pfannstiel, D.C. & Hunter, M.S. (1968). Extending Cooperative Extension education to Mexican-American families - Program methods and evaluation. Abstracted in Science Management Corporation. (1979). (as referenced in Hardee above). (pp. 129-130).

Rossi, P. H., & Freeman, H. E. (1982). Evaluation: A systematic approach. Beverly Hills: Sage.

Seiders, R.W., II. (1972). Dietary adequacy of homemakers participating in Extension's EFNEP in selected Tennessee counties. (Thesis) Abstracted in Kappa Systems, Inc. (1979). (as referenced in Abbott, above). (pp. 199-201).

Further Readings - Comparison Group

Bond, T.H. (1961). An evaluative study of a district Extension program in marketing. Master's Thesis, Michigan State University. Abstracted in Science Management Corporation. (1979). Final report: A review and appraisal of studies of Extension Vol. II. (pp. 71-73). Washington, DC: Author.

Bridges, R. H. (no date). The relative effectiveness of different approaches to training leaders in low income rural communities. Abstracted in Kappa Systems, Inc. (1979). Extension program impact findings from selected studies conducted from 1961 to 1978, Vol. II. (pp. 71-72). Arlington, VA: Author. Appendix to Vol. II, Supporting Abstracts.

Dawson, J. I. (1975). An evaluation of the effectiveness of a leadership development program for rural lay leaders in Lawrence County, Alabama. Abstracted in Kappa Systems, Inc. (1979) (as referenced in Bridges above). (pp. 73-75).

Hatry, H.P., Winnie, R.E., & Fisk, D.M. (1981). Practical program evaluation for state and local government. Washington, DC: Urban Institute.

Neilson, J. (1967). The change agent and the process of change. Abstracted in Science Management Corporation (1979). (as referenced in Bond above). (pp. 26-29).

Prasad, D.C. (1962). A comparison of personal and social development and interest patterns between matched groups of ninth grade 4-H and non-4-H members in ten Wisconsin rural communities. Abstracted in Science Management Corporation. (1979). (as referenced in Bond above). (pp. 231-232).

Rossi, P. H., & Freeman, H. E. (1982). Evaluation: A systematic approach. Beverly Hills: Sage.

Rutledge, L.L. (1971). Selected socio-psychological behavioral changes occurring in boys and girls involved in the Arkansas special 4-H project. Abstracted in Science Management Corporation. (1979). (as referenced in Bond above). (pp. 233-234).

Further Readings - Field Experiment

Bennett, C.F., & Leonard, R. C. (1970). Field experimentation in rural sociology. Rural Sociology, 35, 69-76.

Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi experimental designs for research on teaching. Chicago: Rand McNally.

Fairweather, G.W. (1967). Methods for experimental social innovation. New York: John Wiley & Sons.

Kaufman, R., & Thomas, S. (1980). Evaluation without fear. New York: New Viewpoints.

Rossi, P.H., & Freeman, H.E. (1982). Evaluation: A systematic approach. Beverly Hills: Sage.

Smith, H.W. (1981). Strategies of social research, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Young, R.E. (19[illegible]). The effectiveness of a correspondence course for new 4-H leaders. Doctoral dissertation, the Ohio State University. Abstracted in Science Management Corporation. (1979). Final report: A review and appraisal of studies of Extension Vol. II. Washington, DC: Author.

92

PART III--SUMMARY AND CONCLUSION

CHAPTER 6: SUMMARY

Decisions, Decisions

Like an art collector setting out to choose a work of art, you as a program evaluator, evaluation sponsor or advisor have abundant choices. Reading this resource publication is like browsing through an art gallery. You began in the gallery's main hall and have passed through the corridors displaying each style of art, examining some of the possibilities. You stand, once again, in the main hall. It's now time to begin your reconsideration of types of art in order to select a suitable piece. Remember that you may find no single piece of art to be perfect: the beauty of each is in the eyes of its beholders.

Likewise, there may be no single best design for an evaluation, because evaluation is itself an art (Cronbach, '82). Your choice or recommendation will depend partly on such factors as your administrative, program delivery, or evaluation responsibilities. Your choice or recommendation also depends on budgets, time available, anticipated criticisms of the capacity and rigor of a design, and the political climate. As you plan for a study of program results, select a study design as you would a piece of art, fitting it to the preferences of users and resources within the situation at hand.

Value judgments, concerning which Extension programs are to be evaluated formally and how, operate all along the line. Some of these value judgments are explicit; others are implicit.

Value judgments on selection of study design and facets are best made explicitly, from the start, and continually reviewed. Such value judgments will probably influence the study findings and the ensuing evaluations of a program. Accordingly, we advise study teams to record their premises and value judgments in selecting and implementing a study design and design facets, with a view toward clarity, honesty, and believability.

A Summary of Starting Points

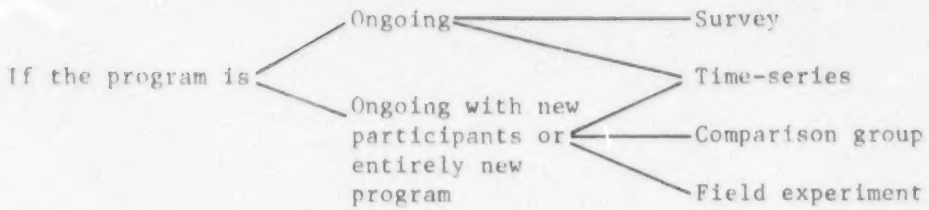
You may now be ready to choose or formulate a design in planning a study of program results, but perhaps the many observations you have made while "browsing" have left you rather overwhelmed. So, at the risk of oversimplifying an individualized process, we offer you seven starting points for selecting a design to study Extension program results. These starting points tend to follow the facets of design discussed earlier.

The seven starting points are expressed graphically below in the form of "if-then" guidelines. In a manner of speaking, if you view your situation as corresponding to "A," then consider options "M or N." Or, if your situation is "B," then consider options "N, Y or Z." In guidelines one and two, the "ifs" refer to Extension programming situations. In guidelines three through six the "ifs" refer to

requirements for intended studies of program results. Guideline seven refers to both program and study requirements.

In each figure below, the "if" situations or conditions are represented in the center column, with the various "then" options presented in the right column. There may be occasional exceptions to the generalizations in these "if-then" guidelines.

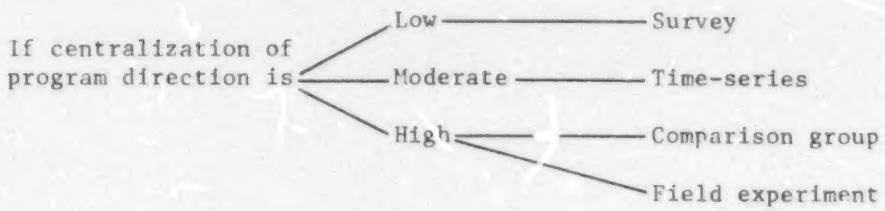
Figure 6.1: Program Status



The decision to examine results of an ongoing program, i.e., one already being implemented, generally necessitates the selection of a survey or time-series design. Measuring results of an ongoing program in which no benchmark data are available requires using a survey design. If benchmark data are available, a time-series design is useful.

The decision to examine results of (a) an ongoing program with clearly defined groups of new participants, or (b) an entirely new program to be implemented allows use of any of the four study designs. Studies of results within these two programming situations can be achieved through the time-series, comparison group and field experiment; these three designs require or allow data to be collected before and after program implementation.

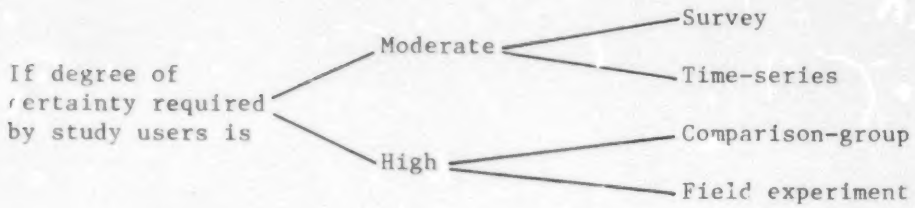
Figure 6.2: Centralization of Program Direction



Studies of the results of programs which are variable over time and by geographic site are generally more feasible through the survey design. Such variability in specific program objectives, audience and delivery methods occurs when Extension programs have decentralized program direction, e.g. county direction of programs is emphasized rather than State direction. The flexibility of the survey design permits it to cover a broad range of program events and consequences. Subjectivist study methods which are frequently employed in the survey design permit aggregating specific program results of a variable nature into more global categories of program results.

Studies of the results of programs with more centralized direction may use the time-series or between-group designs. Greater centralization of direction usally means more standardization of the program across state, district or county lines. And, program standardization aids objectivist methodologies such as highly quantified measurements of changes in clientele performance/status which are typical of time-series, comparison group and field experimental studies. High program standardization also contribute greatly to the formation and use of comparison and/or control groups. Programs which are authorized and funded wholly or primarily from federal or state levels are typically more standardized.

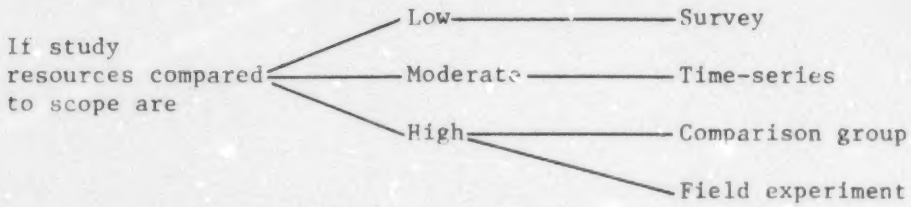
Figure 6.3: Degree of Certainty Needed Regarding Program Results



If a significant evaluation audience has demanding standards for proof that an Extension program caused or contributed to clientele behavior or status, the evaluator should consider a "between-group" design, i.e., a comparison group or field experiment. These designs are particularly effective in proving that changes in people or their environment are due all or in part to an Extension program rather than to other factors. However, the survey and time-series designs are also capable of producing credible findings on Extension program results, depending upon audience for the study, instrumentation, data collection techniques and especially statistical analysis and control. Moreover, the survey may be the only design feasible for studying results of some programs.

Reliability of evidence on results of an Extension program may have implications for the way in which the Extension program is conducted, as well as how it is evaluated. Study designs which provide the higher degrees of scientific certainty regarding an Extension program's (vs. other factors') results also pose constraints on how the program is conducted. For example, a between-group design may affect program timing, program cost per participant and location of program implementation.

Figure 6.4: Study Resources

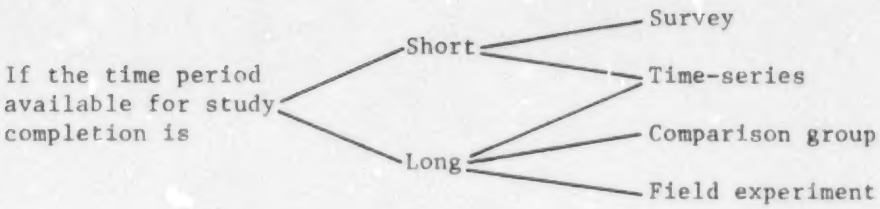


The questions of how many dollars and how much staff time will be needed for each of the study designs under consideration must be addressed. In general, the survey design costs less per person (or group) covered by the study. Surveys based on retrospective data regarding program results may use relatively simple (and thus less costly) statistical treatment of the data. Time-series studies require more frequent data collection than surveys and thus use more resources. Relative to the time-series study, adding data from comparison or control groups for the between-group designs also increases the amount of data to be collected and analyzed.

Concerning scope of programming examined for results, the survey is likely to provide broadest coverage. However, the evidence on program results from a survey is more likely to be based on client perceptions.

Program evaluation expertise is also a resource that may vary by study design selected. We do not assert that some study designs necessarily require more evaluation expertise than others, but rather that study design selected must match professional interests and competencies of staff available.

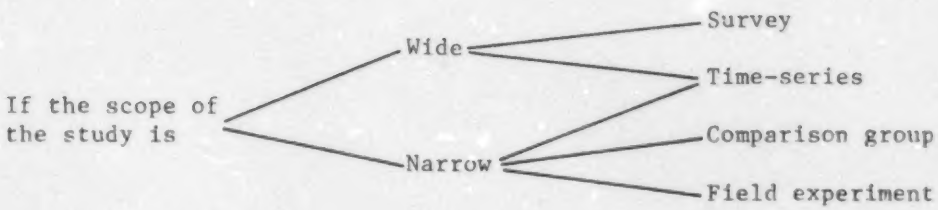
Figure 6.5: Study Time Period



Time period available for the study depends on such factors as how soon documented results are needed, and how long those conducting the study will be available. The survey design lends itself to shorter time periods for study completion. The retrospective nature of the survey makes it unnecessary for the study time period to be as long as the time needed for program effects to occur. This is because, with retrospective data, multiple data collection passes are not employed.

If more time is available for study completion, the time-series, comparison group and field experimental designs are favored. In these three designs it is usually necessary to "wait" for program results to occur so that data may be collected to see if they do in fact occur. The time period needed to conduct a time-series or between group design may be brief if the study is seeks shorter-term rather than longer-term program results. And, if data on clientele regarding the "before" time period are already available, the exercise of these designs may be hastened.

Figure 6.6: Scope of Study

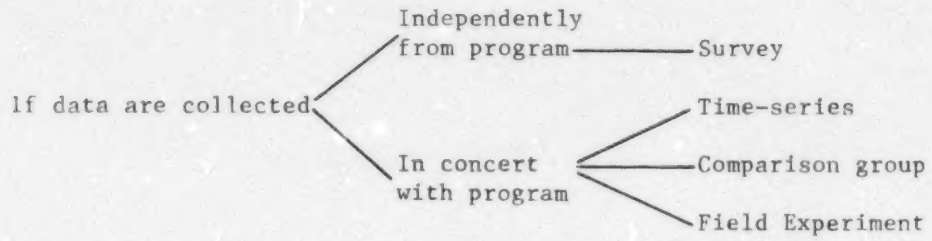


Wide-scope studies, both in terms of geographic area and program content or subject-matter, are easier to achieve through the survey design. The cost per person or group is lower in surveys than in other designs, allowing for a wider-scope study with a given amount of study resources.

The generalization that the survey lends itself to wide-scope studies is applicable particularly to studies that examine a wide subject-matter area. The data collection instrument of a survey can more feasibly cover a broad range of events and consequences. The specific, quantitative comparison typically pursued in the time-series, comparison group and field experiment encourage studying the results of programs with narrower subject-matter content.

Time-series and between-group designs are more manageable for examining results of narrower-scope programs. These designs generally require data to be collected several times. The collection brings the examiner in contact more with delivery of the program being studied. However, it is feasible and may be desirable to conduct time-series and between group studies over wide geographical areas.

Figure 6.7: Coordination of Program Delivery and Data Collection



The survey allows data to be collected independently from the delivery of the program while the other designs generally require data collection to be in concert with program implementation. That is, the timing, personnel and purpose of a survey study are not necessarily closely associated with the timing, personnel and objectives of the program studied. A survey can be conducted long after a program is underway, by individuals not known by the program staff and to measure results not necessarily matching those sought by the program staff.

On the contrary, the time-series, comparison group and field experiment designs must be more closely coordinated with program planning and delivery. Timing, personnel and results to be achieved and investigated are interdependent. Especially with the between group designs, both program planning and how results are to be examined must go hand-in-hand. Where, when, how, and with whom the program will be implemented and not implemented is of mutual concern to both program staffs and evaluation staffs, or to each of these roles if the same Extension personnel are responsible for both program delivery and documentation of program results.

Summary of Text

In Part I (Chapters 1 and 2) we provided the background and approach of this publication, along with an overview of general guidelines and standards for program evaluations. We also described the procedures for obtaining the examples for Volume I of this publication, i.e., the 42 abstracted studies of Extension program results in Volume II. Two social science research firms with experience in assisting federal and state agencies in program evaluation established a pool of 148 positively appraised studies of Extension program results, from the period 1961-1978. Using the procedures and technical accuracy criteria developed by these firms, ES-USDA added five studies from 1979-1982 to the pool, for a total of 153 studies. The studies in Volume II were selected from the above pool of 153 studies.

In Part II (Chapters 3, 4 and 5) four study designs for examining Extension program results were described along with examples of the application of these designs. These examples show feasible ways of undertaking studies according to the four study designs. We systematically identified ten facets of study designs to illustrate the many options open to you in formulating and implementing a study design. We cited examples of these options at work, within each of the four designs, among the studies abstracted in Volume II. Finally, we discussed issues regarding implementing each of the designs.

Here, in Part III, we invited you to reconsider some of the starting points from which you might decide on one or another of the

designs. We acknowledge that you may conclude that some other approach to describing Extension program results suits you better. Perhaps further exploration outside the framework of the present designs may be in your interest. However, if your interest is to develop a study based on quantitative analysis, this publication may serve as a resource.

The starting points from which you might choose one of the four designs involve "if-then" generalizations, to help deal with ideal and practical considerations relating to selection of a study design. These "if-then" guidelines go further than many writings in the field of evaluation in systematically setting forth a "situational approach" to selecting a study design. For example, Campbell and Stanley (1966) confine discussion of the criteria for selection of a design mainly to how much certainty is needed regarding program results.

However, Cronbach (1982) broadens the set of criteria for selecting a study design to include other factors, such as needed scope of a study. Relative to selecting an overall approach to an evaluation, Forest (1980) and Forest and Boyd (1982) include such factors as the organizational level at which information on program results is needed and potential for clientele involvement in study procedures. Patton (1982) relates the situational approach to planning and conducting program evaluations to the standards for program evaluations which are suggested by the Joint Committee on Standards for Educational Evaluation (1981), i.e., utility, feasibility, propriety and accuracy.

References

Campbell, D.T. & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin.

Cronbach, L.J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

Forest, L. (1980). Program evaluation for Extension. Paper presented at the North Central Staff Development Workshop on Evaluation and Accountability. Madison, Wisconsin: University of Wisconsin - Extension.

Forest, L. & Rossing, B.E. (1982). Human values and program evaluation. Journal of Extension, 20, pp. 32-37.

Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluations of education programs, projects and materials. NY: McGraw-Hill.

Patton, M.Q. (1982). Practical evaluation. Beverly Hills, CA: Sage.

CHAPTER 7: IN CONCLUSION

We hope that this publication will assist Extension program leaders and specialists in selecting or advising on the selection of study designs and design facets. We feel that the choice of study design for ascertaining results of Extension programs is strongly relevant to the role of Extension program leaders and specialists.

An evaluation study's quality and relevance (to program staffs' responsibilities for program development, management and accountability) are greatly affected by the choice of study design. We feel that program leaders and specialists have much to contribute to choice of study design without becoming specialists in program evaluation.

Dimensions and Steps of Program Evaluation

In placing this publication in perspective, we have emphasized that evaluation of program results is, so to speak, "only one piece of the evaluation pie;" evaluations which focus on several aspects of Extension programming other than program results should be duly considered in choosing where to expend perennially scarce resources for formal evaluation (Gross, 1977; Longest, 1975). Furthermore, evaluating program results requires expertise far beyond that on study design, including expertise on: managing the study effort; deciding who needs to know what information through involving study users in establishing criteria for evaluating program results; data gathering, data analysis and interpretation; and reporting, dissemination and

utilization of the study findings, conclusions and recommendations (Byrn, 1965; French, 1983; Patton, 1978). However, when considering these other steps in the overall evaluation process, readers should recall that study design is intimately related to every other step. That is, early steps in planning for a study of program results will affect choice or formulation of a study design, which in turn will bear upon the study's findings and utility.

Approaches and Methodologies

The study design matrix and facet analysis in this publication are presented as frameworks to help with study design selection and formulation. We feel that the design options presented can facilitate quantitative study designs that maximize certainty regarding program results, consistent with realistic programming and evaluation situations. The four designs and the ten facets identified are in no way meant to limit design options to be employed in the future. As mentioned previously, we do not intend to imply that qualitative approaches are to be ignored in evaluation of Extension programs. Nonquantitative approaches are simply not within the scope of this publication.

As stated earlier, the studies abstracted in Volume II of this resource do not necessarily represent either all recent, or the best of adequately conducted quantitative studies of Extension program results. We sought variety in types of evidence concerning the influence of Extension programs on clientele behavior or status.

Studies were selected according to different program areas, geographic areas and levels of resources expended.

While all of the abstracted studies in Volume II meet the criteria accepted for technical accuracy, some of the studies have exceptionally high conceptual clarity and exemplary methodology. Several of the studies show creative connections between innovative Extension programming and rigorous evaluation methodology. Generally, the studies seem well designed to have achieved their immediate objectives, that of reporting results of selected Extension programs or, the costs and results of alternate modes of delivery of these programs.

While we intend for this publication to be useful for designing statewide studies of Extension program results, we were able to find only a few such studies as examples. However, we feel most study methodologies exemplified in Volume II can be employed as starting points in designing statewide studies. Although the majority of the abstracts in Volume II focus on one program within a program area, studies of Cooperative Extension program results need not be limited to a single subject matter or program area.

The example studies in Volume II provide background knowledge for designing innovative, but also experienced-based, studies on program results. Readers are encouraged to use the example studies as sources of ideas rather than models to be replicated. We in no way wish to place parameters around future efforts to evaluate Cooperative Extension programs.

Past as Prologue

Finally, we refer to some early perspectives on evaluating results of Extension's educational and information-transfer programs. While formal evaluation of public programs has soared in the last two decades, it began early in Extension. The spirit and philosophy of Extension work has always encouraged systematic study of program effectiveness.

Nearly sixty years ago, the U.S. Department of Agriculture published, "The Effectiveness of Extension in Reaching Rural People" (Wilson, 1926). Wilson asked in this early, four-state study:

--Is Cooperative Extension really reaching large numbers of rural people?

--What better methods of farming and homemaking have been accepted as a result of Extension teaching?

--Which Extension methods have been most effective in obtaining the adoption of these improved practices?

--In the effective conduct of Extension work, how important are land ownership, size of farm distance from the Extension office, membership in the Extension association, contact with Extension workers, and participation in Extension activities?

--What do farmers and farm women think of Extension work now that it has become established in a large number of counties?

While the specific scope of the above questions may appear dated, recast in a more general fashion, they still require answers today. There is indeed prologue in the past, and it deserves our attention.

For example, Wilson's and Gallup's (1955) comprehensive synthesis of studies of Extension program effectiveness, an Extension classic, reported important generalizations on the relation between alternate program delivery methods and clientele use of recommended agricultural and home economics practices.

More broadly, Price (1952) in The Spirit and Philosophy of Extension Work, called for five lines of investigation which appear to be entirely relevant to today's situation. These five lines of investigation concern:

(1) the philosophy of the Extension movement;

(2) the status of those by whom Extension is provided;

(3) the content of Extension programs;

(4) methods for planning, teaching, and evaluating Extension programs;

(5) administration and supervision of Extension programs.

Numerous studies and publications from each preceding decade of Cooperative Extension's history reflect the concern with understanding more about Extension program results. Thus, we carry on a firmly ensconced tradition: the desire to know what effects programs have had, are having, or may have. In a spirit of continuity we offer this publication, and we hope that it is accepted within a philosophy of mutual commitment and sharing.

References

Byrn, D. (Ed.). (1965). Evaluation in Extension. Topeka, KS: H.M. Ives & Sons, Inc.

French, A.M. (1983). A guide to evaluation in the expanded food and nutrition program. Washington, DC: U.S. Government Printing Office.

Gross, J.G. (1977). Evaluation planner for Extension. (Extension Service , Circular 585). Washington, DC: United States Department of Agriculture.

Longest, J. (1975). Designing evaluative research. Journal of Extension, 13, 48-55.

Patton, M.Q. (1978). Utilization-focused evaluation. Beverly Hills: Sage.

Price, M. (1952). Research affecting Extension procedure. In R. K. Bliss, T. B. Symons, M. L. Wilson, G. Gallup, M. J. Reese, and L. M. Schruben (Eds.). The spirit and philosophy of Extension work. Washington, DC: USDA Graduate School and Epsilon Sigma Phi. (p. 203).

Wilson, M.C. (1926). The effectiveness of Extension in reaching rural people. (Departmental Bulletin No. 1384). Washington, DC: United States Department of Agriculture.

Wilson, M.C., & Gallup, G. (1955). Extension teaching methods and other factors that influences adoption of agricultural and home economics practices. (Extension Service, Circular 495). Washington, DC: United States Department of Agriculture.

Recommended Readings

Anderson, S.B., Bell, S., & Murphy, R.T. (1975). Encyclopedia of educational evaluation. San Francisco: Jossey-Bass.

Bennett, C.F. (1979). Analyzing impacts of Extension programs. Washington, DC: U.S.D.A. Science and Education (No. ESC-57S). (rev. ed.).

Campbell, D.T., & Stanley J.C. (1963). Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin.

Cronbach, L.J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

Evaluation Research Society (ERS). (1982). Evaluation research society standards for program evaluation. In P.H. Rossi (Ed.), Standards for evaluation practices (pp. 7-19). San Francisco: Jossey-Bass.

Hopkins, K. D., and Glass, G. V. (1978). Basic statistics for the behavioral sciences. Englewood Cliffs, N.J.: Prentice-Hall.

Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluations of educational programs, projects and materials. New York: McGraw-Hill.

Kaufman, R. & Thomas, S. (1980). Evaluation without fear. NY: New Viewpoints.

Kosecoff, J. and Fink, A. (1983). Evaluation basics: A practitioner's guide. Beverly Hills: Sage.

Mager, R. F. (1972). Goal Analysis. Belmont, CA: Fearson Pitman.

MAGI Educational Services. (1979). Developing effective evaluations for adult education programs: A handbook for administrators and evaluators. Washington, DC: Office of Education.

Miller, R.N. (1979). Evaluative research in rural development: Concepts, methods, issues. Ithaca, NY: Northeast Regional Center for Rural Development.

National Association of State Universities and Land Grant Colleges. (1981). Report of the National Task Force on Extension accountability and evaluation system. Madison: University of Wisconsin-Extension.

Patton, M.Q. (1982). Practical evaluation. Beverly Hills: Sage.

Patton, M.Q. (1980). Qualitative evaluation methods. Beverly Hills: Sage.

Patton, M.Q. (1978). Utilization-focused evaluation. Beverly Hills: Sage.

Smith, H.W. (1980). Strategies of social research: The methodological imagination (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Stufflebeam, D. & Webster, W. (1980). An analysis of alternate approaches to evaluation. Educational evaluation and policy analysis, 2, 5-15.

Sudman, S. (1976). Applied sampling. NY: Academic Press.

Sudman, S., & Bradburn, N. (1982). The art and science of asking questions. Beverly Hills: Sage.

Weiss, C. (1972). Evaluation research: Methods of assessing program effectiveness. Englewood Cliffs, NJ: Prentice-Hall.

Worthen, B.R. and Sanders, J.R. (1973). Educational evaluation theory and practice. Worthington, Ohio: Charles A. Jones.

APPENDICES

Appendix A: Procedures for Abstracting and Appraising Studies

In partial response to a mandate, in the Congressional Food and Agriculture Act of 1977, for an evaluation of the consequences of Cooperative Extension programs, a nationwide search was conducted for analytical studies containing findings on Extension program impacts. Some 450 reports from this search appeared, upon brief inspection, to contain evidence on results from people's participation in Extension programs during the years 1961-1978.

In September 1978, contracts were awarded to Kappa Systems, Inc., and (KSI) Science Management Corporation (SMC) to "review, appraise, and summarize" these 450 studies. An advisory group of Cooperative Extension personnel from each region of the United States was actively involved by Extension Service, USDA, in employing Kappa Systems, Inc., and Science Management Corp, and in oversight of their procedures and report preparation. Two officials of the Budget and Program Analysis office of the U.S. Department of Agriculture also advised ES, USDA, on the contract award to, and procedures used by, the two social research firms. Evaluation and program personnel from several state Extension Services also reviewed the final draft report and suggested revisions in these research contractors' establishment of criteria for technical accuracy of studies of Extension program results. SMC dealt primarily with studies completed during the 1960's and KSI with studies completed during the 1970's.

In the judgement of the contractors approximately 350 of the 450 studies assigned to them actually meet the criteria of (a) contained data on program results rather than just program processes; (b) focussed on programs clearly identified as Extension programs or specifying Extension's contribution to a joint agency

program; and (c) data supplied by program participants or clientele rather than just by Extension staff.

Each study report meeting the minimal qualifications above was condensed into a two-to-four page abstract. Each of the approximately 350 abstracts include a brief description of the program examined, the methodological approach employed by the study, and the findings and any conclusions drawn about the program impact(s) by the study author(s). Thus, program result findings and conclusions are included in the abstracts, rather than study findings and conclusions as a whole.

Appraisals were made as to whether each study's findings and conclusions were "warranted;" i.e., whether there was adequate substantiation according to data presented in the report and research methodology of the author(s). The appraisal criteria included:

a) consistency of study hypotheses and measurements with objectives and structure of the Extension programs studied;

b) validity of measurements and analysis of the data;

c) basis for any inferences regarding results of the program studied

d) basis for any generalization of sample findings to a larger population;

e) primary data collection from clients with adequate size of sample and response rate.

One hundred and forty-eight study reports were found by Kappa Systems, Inc. (1979) and Science Management Corporation (1979) to contain Extension program result findings and conclusions which are warranted relative to evaluative research criteria. The 148 studies are reviewed as a somewhat representative assortment of externally verified examples of program result findings across Extension's program areas.

Forty-nine study reports assigned to KSI for review and appraisal were assigned also to SMC. An analysis of similarity-dissimilarity between the two contractors' reviews and appraisals was based primarily upon the following question: To what extent did both companies agree that the findings and conclusions of each of the 49 reports are warranted?

The two contractors agreed upon 71% of the study reports (35 of the 49 reports) as to whether findings and conclusions on program results were sufficiently warranted by study methodology. Considering only major differences between contractors' judgements on the quality of substantiation of findings and conclusions regarding program results, contractors agreed 92 percent of the time.

The multi-volume reports by each of the two contractors are listed below.

Kappa Systems, Incorporated

Classification of Selected Extension Studies with Impact Findings, Vol. I, Kappa Systems, Inc., Arlington, Virginia.

Extension Program Impact Findings from Selected Studies Conducted from 1961 to 1978, Vol. II, Kappa Systems, Inc., Arlington, Virginia; and Appendix to Volume II, Supporting Abstract.

Guidelines for Improving Extension Program Impact Studies, Vol. III, Kappa Systems, Inc., Arlington, Virginia.

Science Management Corporation

Inventory and Classification of Reports Abstracted, Vol. I, Decision Studies Group, A Division of Science Management Corporation, Washington, D.C.

Final Report: A Review and Appraisal of Studies of Extension Vol. II, Decision Studies Group, A Division of Science Management Corporation, Washington, D.C.

Appendix B: Standards for Appraising Studies

ES-USDA Request for 1979-82 Studies Containing Findings on Extension Program Impacts

This attachment is designed to help the person designated by the state Director or Administrator to find and select a complete study report containing findings on Extension program impacts, for reply to ES-USDA's request dated December 3, 1982. As explained in the memo of request to the state Directors and Administrators, our immediate intent is to add to a pool of example studies from 1961-1978 with examples from 1979-1982. The University of Maryland's Cooperative Extension Service/Department of Agricultural and Extension Education will then select from this pool the most appropriate examples for inclusion in a guidebook for state studies of program impact.

As you search for and select a study to forward to ES-USDA, keep in mind that we need studies which are methodologically adequate relative to standards which were formulated and employed in a past review and appraisal of studies on Extension program effectiveness. We are aware that it may not be possible for some Extension organizations to locate a study of their programming which meets the criteria set forth below.

A full explanation of the processes of review and appraisal used by Science Management Corp (SMC) and Kappa Systems, Inc. (KSI) is presented in Volume III of Kappa System's contractual report, "Guidelines for Improving Extension Program Impact Studies," July 1979. This volume, along with the other volumes from the contractual work to review, appraise and summarize studies from the 1961-1978 were mailed to State Extension Services in September, 1979, as part of the 1978-1980 Congressionally mandated evaluation of economic and social consequences of cooperative extension programs.*

This attachment has been prepared to provide you with enough information on the type of studies we are looking for, so that you do not have to read the Kappa Systems, Inc., materials to respond to our request.

Study reports may represent research from academic departments (including theses) as well as studies sponsored or conducted by your Extension organization. Comparison group, before-after and survey studies will generally be most useful to us.

---

*Volume II by KSI, "Extension Program Impact Findings from Selected Studies conducted from 1961 to 1978" references all 149 studies selected from a larger pool of studies by KSI and SMC as containing impact findings and conclusions with adequate substantiation. Abstracts of these 149 studies may be found in KSI's Appendix to Volume II, and SMC's "Review and Appraisal of Studies of Extension Program Effectiveness," Volume II.

Please select and forward a study of your choice which, in your opinion, meets most of the criteria below:

1. The study report contains findings on Extension program impacts (results). Or, the study report contains findings on clientele's sources or ratings of Extension supplied information, if it does not contain findings on impacts of an Extension program.

   - studies should contain findings within at least one of the following three levels of program impacts:

     (a) learning by Extension program participants (e.g., knowledge, attitude, skill or aspiration change);

     (b) practices or application of learning by program participants; and

     (c) consequences of participants' learning or practices -- economic, social, personal and/or environmental, including benefits, satisfactions, needs or problems.

   - impact data should be collected from (clientele) not only from Extension staff reports.

   - clients' or observers' (e.g., 4-Hers' parents) perceptions of impacts are acceptable as well as objectively measured impacts.

   - studies of clientele's sources or ratings of value of information (including Extension's) may not contain impact findings (see example No. 6 as appended).

2. The study report adequately describes the Extension program being assessed. That is, the report identifies, e.g.,

   - need or rationale for the program

   - program objectives and how they might be accomplished through program activities

   - Extension's particular programming contribution, in the event that a joint agency program is evaluated.

3. The report expresses the purposes of the evaluative study, in relation to:

   - the objectives and structure of program being evaluated

   - rationale for the evaluation--intended importance and utility

   - identification of audiences for whom the evaluation is intended.

4. The study report describes the nature of the "target" population for the program, the extent of clientele participation in the program and the sampling procedure, if any, in obtaining data on program impacts.

   - representativeness of any samples of program participants is described.
   - sample sizes should be greater than 20 (with the exception of units of analysis other than individuals such as firms, organizations and governmental units).

5. The study report establishes a clear link between client outcomes and Extension program delivery, i.e., the report:

   - shows that the Extension program preceded the client outcome
   - shows that degree of clientele outcomes varies with extent of exposure to or involvement in program delivery
   - addresses explicitly or implicitly the extent to which other influences besides Extension could have accounted for the client outcome (clients' self-reported perceptions of degree to which outcomes are due to Extension programs are acceptable).

6. The study report discusses the validity of the measurements or observations of clientele learning, practices and consequences of learning/practices, e.g., the report:

   - shows that the instruments or observation measure variables that are relevant and appropriate
   - shows that the instruments adequately cover the domains of well specified constructs.

7. The study report's findings and conclusions appear to be based on a valid analysis of the data regarding the impacts of the Extension program:

   - logical relationships are established between data sets
   - adequate labelling of tables, charts and graphs
   - clear separation between findings based on data collection and analysis, and general conclusions about the programs' results.

8. The study report provides a comparison if program success or failure is judged, e.g.,

   - program impacts are compared to some established standard or goal, or to impacts of other programs of a similar nature or to absence of a similar program.

Appendix C: Procedures for Obtaining Full Reports of Studies

Cooperative Extension personnel with U.S. Department of Agriculture employee status may obtain photoduplicates of the complete reports of studies abstracted in Volume II, and reports of studies of Extension program results which are referenced in the Further Readings sections of Volume I. These reports are among the holdings in the Extension file at the National Agricultural Library.

Photoduplicates can be obtained from the National Agricultural Library (NAL), Utilization Section, Beltsville, MD, 20705. Telephone requests may be directed to (301) 344-3755 and telecopier requests to (301) 345-6632. It is preferable for Cooperative Extension/USDA staff to submit their requests through the agricultural library at their State land-grant institution.

Requests to the NAL for photoduplicates should contain NAL call numbers assigned to the reports which are desired. Requests bearing call numbers assigned by the National Agricultural Library will be expedited. Call numbers may be obtained through the automated retrieval service, AGRICOLA (Agricultural Online Access). The AGRICOLA computerized bibliographic file is available at 1862 land-grant universities and at most 1890 land-grant institutions.

Full reports of the studies cited in this resource publication may be used at the Beltsville location of the NAL (near the intersection of Route 1 and Interstate Route 95, about 15 miles northeast of Washington, DC) or at the District of Columbia Branch of NAL (Room 1052, South Building, USDA, 14th and Independence Ave., Washington, DC 20250).

Non-USDA employees, including Cooperative Extension personnel without USDA appointment and university students, may obtain photoduplicate reports of the studies cited in this publication through standard interlibrary request procedures. NAL photoduplication charges are: $3.00 for the first 10 pages or fraction thereof from a single publication, and; $2.00 for each additional 10 pages or fraction thereof. Bills will be issued after photocopying is completed.